



IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of: ) Examiner: Saoud, Christine  
Ashkenazi, *et al.* ) Art Unit: 1647  
Application Serial No. 09/905,348 ) Confirmation No: 3826  
Filed: July 13, 2001 ) Attorney's Docket No. 39780-1618 P2C18  
For: **SECRETED AND TRANSMEMBRANE** ) **Customer No. 35489**  
**POLYPEPTIDES AND NUCLEIC** )  
**ACIDS ENCODING THE SAME** )

EXPRESS MAIL LABEL NO. : EV 582 638 100 US  
DATE MAILED: JULY 23, 2007

**ON APPEAL TO THE BOARD OF PATENT APPEALS AND INTERFERENCES**  
**APPELLANTS' BRIEF**

**MAIL STOP APPEAL BRIEF - PATENTS**

Commissioner for Patents  
P.O. Box 1450  
Alexandria, Virginia 1813-1450

Dear Sir:

This Appeal Brief, filed in connection with the above captioned patent application, is responsive to the Final Office Action mailed on October 17, 2006. A Notice of Appeal was filed herein on March 19, 2007. A request for a three-month extension of time is requested herewith. Appellants hereby appeal to the Board of Patent Appeals and Interferences from the final rejection in this case.

The following constitutes the Appellants' Brief on Appeal.

07/26/2007 SSESHE1 00000004 081641 09905348

01 FC:1402 500.00 DA

07/26/2007 SSESHE1 00000004 09905348

02 FC:1253 1020.00 DA

**I. REAL PARTY IN INTEREST**

The real party in interest is Genentech, Inc., South San Francisco, California, by an assignment of the parent application, U.S. Patent Application Serial No. 09/665,350 recorded July 9, 2001, at Reel 011964 and Frame 0181. The present application is a continuation of U.S. Serial No. 09/665,350.

**II. RELATED APPEALS AND INTERFERENCES**

The claims pending in the current application are directed to a polypeptide referred to herein as "PRO232". Although there exist several applications directed to the "gene amplification" utility under Appeal, there are no applications related to PRO232 nucleic acids or antibodies.

**III. STATUS OF CLAIMS**

Claims 44-46 and 49-51 are in this application.

Claims 1-44 and 47-48 have been canceled.

Claims 44-46 and 49-51 stand rejected and Appellants appeal the rejection of these claims.

A copy of the rejected claims in the present Appeal is provided in Section IX.

**IV. STATUS OF AMENDMENTS**

A summary of the prosecution history for this case is as follows:

Previously, in response to a Final Office Action mailed on October 21, 2005, a Notice of Appeal was filed on February 21, 2006 and an Appeal Brief was filed on April 17, 2006. An RCE Response with additional references and affidavits supporting Appellants' arguments was filed on July 27, 2006. A Final Office Action was mailed on October 17, 2006, and a Notice of Appeal was filed on March 19, 2007.

No amendments were submitted after the final rejection mailed October 17, 2006. All previous amendments have been entered.

**V. SUMMARY OF CLAIMED SUBJECT MATTER**

The invention claimed in the present application is related to an isolated polypeptide comprising the amino acid sequence of the polypeptide of SEQ ID NO: 18, referred to in the



present application as "PRO232". The PRO232 gene was shown for the first time in the present application to be significantly amplified in human lung cancers as compared to normal, non-cancerous human tissue controls (Example 92). This feature is specifically recited in Claim 124, and carried by all claims dependent from claim 44. In addition, the invention also claims the amino acid sequence of the polypeptide of SEQ ID NO: 18, lacking its associated signal-peptide; or the amino acid sequence of the polypeptide encoded by the full-length coding sequence of the cDNA deposited under ATCC accession number 209250 (Claim 44-46 and 49). The invention is further directed to a chimeric polypeptide comprising one of the above polypeptides fused to a heterologous polypeptide (Claim 50), and to a chimeric polypeptide wherein the heterologous polypeptide is an epitope tag or an Fc region of an immunoglobulin (Claim 51). The preparation of chimeric PRO polypeptides (Claims 50 and 51), including those wherein the heterologous polypeptide is an epitope tag or an Fc region of an immunoglobulin, is set forth in the specification at page 74, lines 23 to page 75, line 5. Examples 53-56, pages 192-199, describe the expression of PRO polypeptides in various host cells, including *E. coli*, mammalian cells, yeast and Baculovirus-infected insect cells.

The amino acid sequence of the "PRO232" polypeptide and the nucleic acid sequence encoding this polypeptide (referred to in the present application as "DNA34435-1140") are shown in the present specification as SEQ ID NOs: 18 and 17, respectively, and in Figures 9 and 8, described on page 59, lines 4-7. The full-length PRO232 polypeptide having the amino acid sequence of SEQ ID NO:18 is described in the specification at, for example, on page 4, pages 3-4 and page 100, page 131, line 9 to 16 and the isolation of cDNA clones encoding PRO232 of SEQ ID NO:18 is described in Example 4, page 149-150 of the specification. The specification discloses that the PRO232 polypeptide possess significant sequence homology to cell surface stem cell antigen (35% sequence identity with a stem cell surface antigen from *Gallus gallus*) and may play a role in cell proliferation and/or differentiation. (see for example, page 4 and Example 4, line 14-15).

Finally, Example 92, in the specification at page 222, line 26, to page 235, line 3, sets forth a 'Gene Amplification assay' which shows that the PRO232 gene is amplified in the genome of certain human lung cancers (see Table 9, pages 230-234). The profiles of various primary lung and colon tumors used for screening the PRO polypeptide compounds of the

invention in the gene amplification assay are summarized on Table 8, page 227 of the specification.

## **VI. GROUNDINGS OF REJECTION TO BE REVIEWED ON APPEAL**

1. Whether instant Claims 44-46 and 49-51 satisfy the utility/enableness requirement under 35 U.S.C. §§101/112, first paragraph.
2. Whether Claims 44-46 and 49-51 are entitled to the priority date of U.S. Provisional Application 60/059121, filed September 17, 1997.
3. Whether Claims 44-46 and 49-51 are anticipated under 35 U.S.C. §102(b) by Rosenthal et al., DE19818619-A1 (October 1999).

## **VII. ARGUMENTS**

### **Summary of the Arguments**

#### **Issue 1: Utility**

As a preliminary matter, Appellants note that the Examiner refers to the microarray assay in several instances in several Office Actions, and cites references like Lilley *et al.*, King *et al.*, Lee *et al.*, Wildsmith *et al.*, Chen *et al.* etc. which use and analyze the microarray assay. In addition, the Examiner rejects the use of the “universal control” in the instant application. Appellants point out that neither the “universal control” nor the microarray were used in the claimed invention, hence rejections directed either to the limitations of the use of the microarray assay, or to the universal control, are moot. On the other hand, Appellants rely upon the gene amplification data of the PRO232 gene for patentable utility of the PRO232 polypeptides in the present application, and controls defined in Example 92 of the instant specification, and would like to distinguish between the gene amplification and the microarray assay. The gene amplification assay measures the level at which a certain gene (i.e. DNA) is amplified in the genome, whereas the microarray assay measures the level of expression of a mRNA encoding for a certain polypeptide in a sample. Throughout prosecution, the Examiner fails to distinguish between these two techniques, but Appellants submit that the two assays, although similar, are not the same.

Appellants would like to bring to the Examiner’s attention a recent decision in a microarray case by the Board of Patent Appeals and Interferences (Decision on Appeal No.

2006-1469). In its decision, the Board reversed the utility rejection, acknowledging that “there is a strong correlation between mRNA levels and protein expression, and the Examiner has not presented any evidence specific to the PRO1866 polypeptide to refute that.” (Page 9).

Appellants submit that, in the instant application, the Examiner has likewise not presented any evidence specific to the PRO232 polypeptide to refute Applicant’s assertion of a correlation between DNA levels, mRNA levels and protein expression. Appellants add that they discuss or analyze the microarray assay only in response to the Examiner’s citations and in response to correlation of mRNA and protein levels. This is not to be construed as an admission that all that applies to the microarray assay, applies to the gene amplification assay as well.

Appellants rely upon the gene amplification data of the PRO232 gene for patentable utility of the PRO232 polypeptides. The specification discloses that the gene encoding PRO232 showed significant amplification, ranging from 2.056-fold to 5.28-fold, in five lung tumors or 2.00-fold to 5.32-fold in seven colon tumors.

Appellants have submitted, in their Response filed July 25, 2005, a Declaration by Dr. Audrey Goddard, which explains that a gene identified as being amplified at least 2-fold by the disclosed gene amplification assay in a tumor sample relative to a normal sample is useful as a marker for the diagnosis of cancer, and for monitoring cancer development and/or for measuring the efficacy of cancer therapy. Therefore, such a gene is useful as a marker for the diagnosis of lung or colon cancer, and for monitoring cancer development and/or for measuring the efficacy of cancer therapy.

Appellants have also submitted, in their Responses filed August 9, 2004, and July 27, 2006, ample evidence to show that, in general, if a gene is amplified in cancer it is more likely than not that the encoded protein will be expressed at an elevated level. For instance, the articles by Orntoft *et al.*, Hyman *et al.*, and Pollack *et al.* collectively teach that in general, gene amplification increases mRNA expression. Second, the Declarations of Dr. Paul Polakis: (Polakis I and II), shows that, in general, there is a correlation between mRNA levels and polypeptide levels. Third, Appellants further submit that even if there were no correlation between gene amplification and increased mRNA/protein expression, (which Appellants expressly do not concede to), a polypeptide encoded by a gene that is amplified in cancer would still have a specific, substantial, and credible utility. Appellants submit that, as evidenced by the

Ashkenazi Declaration and the teachings of Hanna and Mornin (both made of record in Appellants' Responses filed December 10, 2003 and August 9, 2004), simultaneous testing of gene amplification and gene product over-expression enables more accurate tumor classification, even if the gene-product, the protein, is not over-expressed. This leads to better determination of a suitable therapy for the tumor, as demonstrated by a real-world example of the breast cancer marker HER-2/neu.

Appellants further note that the sale of gene expression chips to measure mRNA levels is a highly successful business, with a company such as Affymetrix recording 168.3 million dollars in sales of their GeneChip arrays in 2004. Clearly, the research community believes that the information obtained from these chips is useful (*i.e.*, that it is more likely than not informative of the protein level). Therefore, as a general rule, one skilled in the art would find it more likely than not that PRO232 polypeptides are useful as a diagnostic tools for detecting lung or colon tumors.

The Examiner relies on the teachings of Pennica *et al.*, Konopka *et al.*, Haynes *et al.*, Hu *et al.*, Lian *et al.*, Fessler *et al.*, Gygi *et al.*, Lilley *et al.*, Lee *et al.*, King *et al.*, Wildsmith *et al.*, Nagaraja *et al.*, Sagynaliev *et al.*, Waghray *et al.*, Madoz-Gurpide *et al.*, Feroze-Merzoug *et al.*, Bustin *et al.*, Saito-Hisaminato *et al.*, to allege that there is no correlation between increased gene amplification and protein levels.

Appellants respectfully disagree and submit that the teachings within these cited references do not conclusively establish a *prima facie* case for lack of utility because the references are, either not contrary to the Appellants' arguments, or, actually lend support to the Appellants' position, or, are not applicable to the present application for various reasons, as discussed in detail below. On the other hand, Appellants submit that while the literature indicates that some references demonstrate a positive correlation between DNA levels, mRNA expression and protein levels, and some show no correlation, in general, there are more cases in literature that show a positive correlation than not.

Taken together, although there are some examples in the scientific art that do not fit within the central dogma of molecular biology that there is generally a positive correlation between DNA, mRNA, and polypeptide levels, in general, in the majority of amplified genes, as exemplified by the teachings of Orntoft *et al.*, Hyman *et al.*, Pollack *et al.*, the two Polakis

Declarations, the art overwhelmingly show that gene amplification influences gene expression at the mRNA and protein levels. The widespread, art accepted use of information obtained from array chips for detecting diagnostic markers lend further support that in general, one of skill in the art would reasonably expect in this instance, based on the amplification data for the PRO232 gene, that the PRO232 polypeptide is concomitantly overexpressed and has utility in the diagnosis of lung or colon cancer or for individuals at risk for developing lung or colon cancer.

Accordingly, Appellants submit that when the proper legal standard is applied, one should reach the conclusion that the present application discloses at least one patentable utility for the claimed PRO232 polypeptides. Accordingly, one of ordinary skill in the art would also understand how to make and use the recited polypeptides for the diagnosis of lung or colon cancer without any undue experimentation.

### **Issue 2: Priority**

The instant application has not been granted the earlier priority date of U.S. Provisional Application 60/059121, filed September 17, 1997 on the grounds that the 60/059121 application fails to provide a utility and lacks an enabling disclosure for the claimed invention under 35 U.S.C. §§101/112, first paragraph.

Appellants submit that, for the same reasons discussed above under Issue 1, U.S. Provisional Application 60/059121 also satisfies the utility requirements. Therefore, Appellants should be entitled to the priority date of **September 17, 1997**.

### **Issue 3: Anticipation by Rosenthal *et al.***

As discussed above under Issue 2, the present application should be entitled to the earlier filing date of **September 17, 1997** and therefore, Rosenthal *et al.*, DE19818619-A1, dated October 1999, is not prior art. Thus the instant claims are not anticipated by Rosenthal *et al.*

These arguments are all discussed in further detail below under the appropriate headings.

### **Response to Rejections**

#### **ISSUE 1: The Instant Claims 44-46 and 49-51 Satisfy the Utility Requirement under 35 U.S.C. §101/§112, First Paragraph based on the results of the gene amplification assay**

The sole basis for the Examiner's rejection of Claims 44-46 and 49-51 under this section is that the data presented in the instant Application, allegedly, does not satisfy the requirements

of 35 U.S.C. §§101/112, first paragraph. Appellants strongly disagree for the reasons discussed below.

**A. The Legal Standard For Utility Under 35 U.S.C. §101**

According to 35 U.S.C. §101:

Whoever invents or discovers any new and *useful* process, machine, manufacture, or composition of matter, or any new and *useful* improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title. (Emphasis added.)

In interpreting the utility requirement, in *Brenner v. Manson*<sup>1</sup> the Supreme Court held that the *quid pro quo* contemplated by the U.S. Constitution between the public interest and the interest of the inventors required that a patent applicant disclose a “substantial utility” for his or her invention, i.e. a utility “where specific benefit exists in currently available form.”<sup>2</sup> The Court concluded that “a patent is not a hunting license. It is not a reward for the search, but compensation for its successful conclusion. A patent system must be related to the world of commerce rather than the realm of philosophy.”<sup>3</sup>

Later, in *Nelson v. Bowler*<sup>4</sup> the C.C.P.A. acknowledged that tests evidencing pharmacological activity of a compound may establish practical utility, even though they may not establish a specific therapeutic use. The court held that “since it is crucial to provide researchers with an incentive to disclose pharmaceutical activities in as many compounds as possible, we conclude adequate proof of any such activity constitutes a showing of practical utility.”<sup>5</sup>

---

<sup>1</sup> *Brenner v. Manson*, 383 U.S. 519, 148 U.S.P.Q. (BNA) 689 (1966).

<sup>2</sup> *Id.* at 534, 148 U.S.P.Q. (BNA) at 695.

<sup>3</sup> *Id.* at 536, 148 U.S.P.Q. (BNA) at 696.

<sup>4</sup> *Nelson v. Bowler*, 626 F.2d 853, 206 U.S.P.Q. (BNA) 881 (C.C.P.A. 1980).

<sup>5</sup> *Id.* at 856, 206 U.S.P.Q. (BNA) at 883.

In *Cross v. Iizuka*<sup>6</sup> the C.A.F.C. reaffirmed *Nelson*, and added that *in vitro* results might be sufficient to support practical utility, explaining that “*in vitro* testing, in general, is relatively less complex, less time consuming, and less expensive than *in vivo* testing. Moreover, *in vitro* results with the particular pharmacological activity are generally predictive of *in vivo* test results, i.e. there is a reasonable correlation there between.”<sup>7</sup> The court perceived “No insurmountable difficulty” in finding that, under appropriate circumstances, “*in vitro* testing, may establish a practical utility.”<sup>8</sup>

The case law has also clearly established that Appellants’ statements of utility are usually sufficient, unless such statement of utility is unbelievable on its face.<sup>9</sup> The PTO has the initial burden to prove that Appellants’ claims of usefulness are not believable on their face.<sup>10</sup> In general, an Applicant’s assertion of utility creates a presumption of utility that will be sufficient to satisfy the utility requirement of 35 U.S.C. §101, “unless there is a reason for one skilled in the art to question the objective truth of the statement of utility or its scope.”<sup>11, 12</sup>

Compliance with 35 U.S.C. §101 is a question of fact.<sup>13</sup> The evidentiary standard to be used throughout *ex parte* examination in setting forth a rejection is a preponderance of the totality of the evidence under consideration.<sup>14</sup> Thus, to overcome the presumption of truth that

---

<sup>6</sup> *Cross v. Iizuka*, 753 F.2d 1047, 224 U.S.P.Q. (BNA) 739 (Fed. Cir. 1985).

<sup>7</sup> *Id.* at 1050, 224 U.S.P.Q. (BNA) at 747.

<sup>8</sup> *Id.*

<sup>9</sup> *In re Gazave*, 379 F.2d 973, 154 U.S.P.Q. (BNA) 92 (C.C.P.A. 1967).

<sup>10</sup> *Ibid.*

<sup>11</sup> *In re Langer*, 503 F.2d 1380, 1391, 183 U.S.P.Q. (BNA) 288, 297 (C.C.P.A. 1974).

<sup>12</sup> See also *In re Jolles*, 628 F.2d 1322, 206 USPQ 885 (C.C.P.A. 1980); *In re Irons*, 340 F.2d 974, 144 USPQ 351 (1965); *In re Sichert*, 566 F.2d 1154, 1159, 196 USPQ 209, 212-13 (C.C.P.A. 1977).

<sup>13</sup> *Raytheon v. Roper*, 724 F.2d 951, 956, 220 U.S.P.Q. (BNA) 592, 596 (Fed. Cir. 1983) cert. denied, 469 US 835 (1984).

<sup>14</sup> *In re Oetiker*, 977 F.2d 1443, 1445, 24 U.S.P.Q.2d (BNA) 1443, 1444 (Fed. Cir. 1992).

an assertion of utility by the applicant enjoys, the Examiner must establish that it is more likely than not that one of ordinary skill in the art would doubt the truth of the statement of utility. Only after the Examiner made a proper *prima facie* showing of lack of utility, does the burden of rebuttal shift to the applicant. The issue will then be decided on the totality of evidence.

The well established case law is clearly reflected in the Utility Examination Guidelines (“Utility Guidelines”)<sup>15</sup>, which acknowledge that an invention complies with the utility requirement of 35 U.S.C. §101, if it has at least one asserted “specific, substantial, and credible utility” or a “well-established utility.” Under the Utility Guidelines, a utility is “specific” when it is particular to the subject matter claimed. For example, it is generally not enough to state that a nucleic acid is useful as a diagnostic without also identifying the conditions that are to be diagnosed.

In explaining the “substantial utility” standard, M.P.E.P. §2107.01 cautions, however, that Office personnel must be careful not to interpret the phrase “immediate benefit to the public” or similar formulations used in certain court decisions to mean that products or services based on the claimed invention must be “currently available” to the public in order to satisfy the utility requirement. “Rather, any reasonable use that an applicant has identified for the invention that can be viewed as providing a public benefit should be accepted as sufficient, at least with regard to defining a ‘substantial’ utility.”<sup>16</sup> Indeed, the Guidelines for Examination of Applications for Compliance With the Utility Requirement,<sup>17</sup> gives the following instruction to patent examiners: “If the applicant has asserted that the claimed invention is useful for any particular practical purpose . . . and the assertion would be considered credible by a person of ordinary skill in the art, do not impose a rejection based on lack of utility.”

**B. Proper Application of the Legal Standard**

Appellants respectfully submit that the data presented in Example 92 starting on page 222 of the priority application and the cumulative evidence of record, which underlies the current

---

<sup>15</sup> 66 Fed. Reg. 1092 (2001).

<sup>16</sup> M.P.E.P. §2107.01.

<sup>17</sup> M.P.E.P. §2107 II (B)(1).



dispute, indeed support a “specific, substantial and credible” asserted utility for the presently claimed invention.

Patentable utility for the PRO232 polypeptides is based upon the gene amplification data for the gene encoding the PRO232 polypeptide. Example 92 describes the results obtained using a very well-known and routinely employed polymerase chain reaction (PCR)-based assay, the TaqMan™ PCR assay, also referred to herein as the gene amplification assay. This assay allows one to quantitatively measure the level of gene amplification in a given sample, say, a tumor extract, or a cell line. It was well known in the art at the time the invention was made that gene amplification is an essential mechanism for oncogene activation. Appellants isolated genomic DNA from a variety of primary cancers and cancer cell lines that are listed in Table 9 (pages 222 onwards of the specification), including primary lung and colon cancers of the type and stage indicated in Table 8 (page 227). The tumor samples were tested in triplicates with Taqman™ primers and with internal controls, beta-actin and GAPDH in order to quantitatively compare DNA levels between samples (page 229). As a negative control, DNA was isolated from the cells of ten normal healthy individuals, which was pooled and used as a control (page 222, lines 28-29). The results of TaqMan™ PCR are reported in  $\Delta$ Ct units, as explained in the passage on page 222, lines 37-39. One unit corresponds to one PCR cycle or approximately a 2-fold amplification, relative to control, two units correspond to 4-fold, 3 units to 8-fold amplification and so on. Using this PCR-based assay, Appellants showed that the gene encoding for PRO232 was amplified, that is, it showed approximately 1.04-2.40  $\Delta$ Ct units for five lung tumors and 1.00-2.41  $\Delta$ Ct units for seven colon tumors, which corresponds to  $2^{1.04}$ - $2^{2.40}$ - fold amplification in lung or to  $2^{1.00}$ - $2^{2.41}$ - fold amplification in colon tumors; that is **2.056-fold to 5.28-fold in five lung tumors** or **2.00-fold to 5.32-fold in seven colon tumors**, which would be considered significant and credible by one skilled in the art. Therefore, the PRO232 gene and the PRO232 polypeptide are important diagnostic markers to identify such malignant lung or colon cancers.

**A prima facie case of lack of utility has not been established**

As discussed above, the increase in DNA copy number for the PRO232 gene is significant. Further, the evidentiary standard to be used throughout *ex parte* examination of a patent application is a preponderance of the totality of the evidence under consideration. Thus, to overcome the presumption of truth that an assertion of utility by the applicant enjoys; the

Examiner must establish that it is more likely than not that one of ordinary skill in the art would doubt the truth of the statement of utility.

Accordingly, it is not a legal requirement to establish a necessary correlation between an increase in the copy number of the DNA and protein expression levels that would correlate to the disease state or that it is imperative to find evidence that DNA amplification is "necessarily" or "always" associated with overexpression of the gene product. Appellants respectfully submit that when the proper evidentiary standard is applied, a correlation must be acknowledged. Only after the Examiner has made a proper *prima facie* showing of lack of utility, does the burden of rebuttal shift to the Applicant.

Previously, the Examiner has indicated based on references Pennica *et al.*, Konopka *et al.* and Haynes *et al.* to show that gene amplification data cannot reliably predict protein levels. Appellants have argued the references in great detail throughout prosecution and these arguments are incorporated by reference herein for brevity. Appellants summarize the rejections and the arguments submitted below.

The teachings of Pennica *et al.* are specific to *WISP* genes, a specific class of closely related molecules. Pennica *et al.* showed that there was good correlation between DNA and mRNA expression levels for the *WISP-1* gene, but not for *WISP-2* and *WISP-3* genes. *WISPs* 1-3 have no structural relationship to the PRO232 polypeptides of the present application. The apparent finding that for two out of three specific molecules, that are related to each other but have no relationship to PRO232, that there was no correlation between gene amplification and the level of mRNA/protein expression does not establish, in general, that it is more likely than not that such correlation does not exist, and has no bearing whatsoever on determining the question whether such correlation is likely to exist between PRO232 gene amplification and mRNA/protein expression levels. As discussed above, the standard is not absolute certainty. Pennica *et al.* has no teaching whatsoever about the correlation of gene amplification and protein expression for genes in general, or PRO232 or related molecules in particular.

Similarly, in Konopka *et al.*, the Examiner has generalized a very specific result disclosed by Konopka *et al.* to cover all genes. Konopka *et al.* actually state that "[p]rotein expression is not related to amplification of the *abl* gene but to variation in the level of *bcr-abl* mRNA produced from a single Ph<sup>1</sup> template." (See Konopka *et al.*., Abstract, emphasis added). The

paper does not teach anything whatsoever about the correlation of protein expression and gene amplification in general, and provides no basis for the generalization that apparently underlies the present rejection. The statement of Konopka *et al.* that “[p]rotein expression is not related to amplification of the *abl* gene . . .” is not sufficient to establish a *prima facie* case of lack of utility. Therefore, the combined teachings of Pennica *et al.* and Konopka *et al.* are not directed towards genes in general but to a single gene or genes within a single family and thus, their teachings cannot support a general conclusion regarding a correlation between gene amplification and mRNA or protein levels. In addition, the *abl* gene has no structural relationship to the PRO232 gene of the present application and thus, Konopka *et al.* provides no information of specific relevance to the question whether for PRO232 there is a reasonable expectation that correlation between gene amplification and mRNA/protein expression levels is likely to exist.

The Examiner also cited Haynes *et al.* to show that transcript levels and protein levels do not correlate. However, Appellants had shown that Haynes themselves admit that “there was a general trend, although no strong correlation between protein [expression] and transcript levels” (see Figure 1 and page 1863, paragraph 2.1, last line). Therefore, when the proper legal standard is used, Haynes clearly supports the Appellants’ position that in general, a positive correlation exists between mRNA and protein expression levels. Since accurate prediction is not the standard, a *prima facie* case of lack of utility has not been met based on the cited references Pennica *et al.*, Konopka *et al.* and Haynes *et al.* Appellants respectfully submit that, contrary to the Examiner’s assertion, none of the cited reference conclusively establish a *prima facie* case for lack of utility for the PRO232 molecule.

Appellants have already discussed references Hu *et al.*, Chen *et al.*, Lian *et al.*, Fessler *et al.*, in great detail in their previous responses (at least see Appeal brief filed April 17, 2006), and **these arguments are hereby incorporated by reference for brevity.**

Briefly, the analytical methods utilized by Hu *et al.* have certain statistical drawbacks, as the authors themselves admit, and Hu *et al.*’s conclusions only apply to a specific type of breast tumor (estrogen receptor (ER)-positive breast tumor) and cannot be generalized to breast cancer genes in general, let alone to cancer genes in general.

Regarding Lian *et al.*, Appellants respectfully submit that Lian *et al.* only teach that protein expression may not correlate mRNA level in differentiating myeloid cells, and not about genes in general. In fact, the authors themselves admit that there were a number of problems with their data. For instance, at page 520 of this article, the authors explicitly express their concerns regarding the methods they utilized and the interpretation of their data stating that "[t]hese data must be considered with several caveats: membrane and other hydrophobic proteins and very basic proteins are not well displayed by the standard 2DE approach, and proteins presented at low level will be missed. In addition, to simplify MS analysis, we used a Coomassie dye stain rather than silver to visualize proteins, and this decreased the sensitivity of detection of minor proteins." (Emphasis added). Besides, Lian *et al.*'s conclusions are based on the Coomassie dye staining method, which is not a very sensitive method of measuring protein. Similarly, in Fessler *et al.*, examined lipopolysaccharide-activated neutrophilins, in response to LPS stimulation. Fessler *et al.* also used the Coomassie Blue dye staining method, and concede that it is known to have a limited protein binding range and a non-linear curve for protein detection. Protein identification in their study was also done using two-dimensional PAGE, which is, by their own admission, limited only to well-resolved regions of the gel, and therefore, may have performed less well with hydrophobic and high molecular weight proteins (see page 31301, col. 1).

#### Gygi *et al.*

The Examiner has cited new reference Gygi *et al.* in the Final Office Action of October 17, 2006 addressing the correlation between mRNA and protein levels.

Appellants submit that Gygi *et al.* never indicate that the correlation between mRNA and protein levels does not exist. Gygi *et al.* only state that the correlation may not be sufficient to **accurately** predict the protein level from the level of the corresponding mRNA transcript (see page 1270, Abstract). Gygi *et al.* may teach that protein levels cannot be "predicted" from mRNA levels in the sense that the exact numerical amounts of protein present in a tissue cannot be determined based upon mRNA levels. But Appellants respectfully submit that the PTO's emphasis on the need to "accurately predict" protein levels based on mRNA levels misses the point. The asserted utility for the claimed polypeptides is in the diagnosis of cancer. What is relevant to use as a cancer diagnostic is relative levels of gene or protein expression, not absolute

values, that is, that the gene or protein is differentially expressed in tumors as compared to normal tissues.

Moreover, contrary to the Examiner's statement, the Gygi data indicate a **general trend** of correlation between protein [expression] and transcript levels (Emphasis added). For example, as shown in Figure 5, the mRNA abundance of **250-300** copies/cell correlates with the protein abundance of **500-1000** x 10<sup>3</sup> copies/cell. The mRNA abundance of **100-200** copies/cell correlates with the protein abundance of **250-500** x 10<sup>3</sup> copies/cell (emphasis added). Therefore, high levels of mRNA **generally** correlate with high levels of proteins. In fact, most data points in Figure 5 did not deviate or scatter away from the general trend of correlation. Thus, the Gygi data meets the "more likely than not standard" and shows that a positive correlation exists between mRNA and protein.

In summary, Hu *et al.*, Lian *et al.*, Fessler *et al.*, Gygi *et al.* do not conclusively teach that, in general, protein levels cannot be accurately predicted from mRNA/gene amplification levels. These authors concede that either due to insensitive protein detection methods or due to their methodologies utilized in their protocols, some protein species may have been under-represented over others. Therefore, the teachings of these references cannot be relied upon to establish a *prima facie* showing of lack of utility. On the other hand, as noted even in Haynes *et al.* and Gygi *et al.* most genes showed a positive correlation between increased gene amplification, mRNA and translated protein.

The Examiner asserts that, of the references cited by Appellants in their IDS filed July 27, 2006, "only Godbout *et al.* is pertinent."

Appellants respectfully submit that, as discussed in the RCE response of July 27, 2006, there are several other references; for instance, Bea *et al.* who investigated gene amplification, mRNA expression, and protein expression of the putative oncogene BMI-1 in human lymphoma samples, and support Appellants' assertion that gene amplification is correlated with both increased mRNA and protein expression.

**Godbout *et al.* (Item 46 of Evidence List)**

Regarding the Godbout reference, the Examiner alleges that the instant specification does not teach structure/function analysis; the Examiner says that she finds no reason to suspect if PRO232 can confer selective advantage to a cell. The Examiner questions whether the level of

genomic amplification of DDX1 gene is comparable to that of PRO232 (page 5 of Final Office Action of October 17, 2006).

Appellants respectfully submit that it was never claimed that PRO232 is similar in any way to the DDX1 gene of Godbout *et al.*, they never claimed PRO232 was an RNA helicase or that it confers selective advantage to cell survival; on the other hand, the Godbout reference was submitted to show good correlation between protein levels based upon genomic DNA amplification, which the Examiner clearly agrees with. Moreover, selective advantage to cell survival is not the only mechanism by which genes impact cancer and structure/function data, which the Examiner requests, is not a requirement for the utility requirement. Hence a *prima facie* case has not been established and this rejection is improper.

**Saito- Hisaminato *et al.* and Bustin *et al.***

The Examiner refers to the microarray assay and the “universal control” and indicates that proper controls are needed for the microarray assay. The Examiner rejects the use of the universal control based on the teachings within Saito- Hisaminato *et al.* (Pages 18-19 of the instant Final Office action). The Examiner further cites Bustin *et al.* without any explanations for the citation.

As discussed above under Section VII, Appellants respectfully submit that they rely upon the gene amplification data of the PRO232 gene for patentable utility of the PRO232 polypeptides in the present application, not the microarray assay. Appellants would like to distinguish between the gene amplification and the microarray assay. The gene amplification assay measures the level at which a certain gene (i.e. DNA) is amplified in the genome, whereas the microarray assay measures the level of expression of a mRNA encoding for a certain polypeptide in a sample. Throughout prosecution, the Examiner fails to distinguish between these two techniques, but Appellants submit that the two assays, although similar, are not the same. Any rejection directed to the limitations of the microarray assay or to the universal control are moot and are not addressed here. Only relevance of microarray array results to correlate mRNA to protein levels is discussed.

No “universal control” was used in the instantly claimed invention. Since “Saito- Hisaminato” primarily discusses the universal control, its teachings do not apply to the instant invention in any manner and is moot.

The Examiner has not provided reasons for why Bustin *et al.*, which uses and analyzes microarray technology, was cited. Without conceding to the propriety of this rejection, Appellants note that Bustin *et al.* do not dispute the use of microarrays to obtain biologically relevant data, noting that “several microarray experiments have generated clinically relevant quantitative gene profiles” (page 271, col. 1). In fact, Bustin *et al.* conclude that “[m]icroarrays are already having a major impact on cancer biology, pharmacology and drug development” and state that “the major limiting factor in their further application is the current lack of data comparability, which is essential for appropriate comparisons between different arrays.”

Accordingly, a *prima facie* case has not been established based on Saito- Hisaminato *et al.* and Bustin *et al.* and this rejection is improper.

**Lilley *et al.*, King *et al.*, Lee *et al.* and Wildsmith *et al.***

First of all, the references Lilley *et al.*, King *et al.*, Lee *et al.* and Wildsmith *et al.* all utilize and analyze the microarray assay. The Examiner cites the Lee *et al.* reference allege a limitation that requires replication of microarray assays (see page 14, line 3 of Final Office action). The Examiner cites King *et al.* to allege a limitation that microarray has variability for high- or medium- abundance mRNAs (see page 14, lines 12-13 of Final Office action). The Examiner further indicates, based on Lilley *et al.*, Wildsmith *et al.*, and King *et al.*, that the state of the art demonstrates that correlation between mRNA and protein abundance “cannot be accurately predicted” (page 12 of Final Office action).

Again, as discussed above, without conceding to the propriety of the rejections stated above, Appellants submit that the rejections directed to the limitations of microarray assay are moot and are not addressed here, since Appellants rely upon the gene amplification data of the PRO232 gene for patentable utility of the PRO232 polypeptides and not the microarray assay.

Appellants add that it is not a legal requirement for utility, to establish a necessary correlation between an increase in the mRNA level and protein expression levels; that is, the utility standard is not absolute certainty or it is not necessary to show that changes in transcript level should always result in corresponding changes in protein amount or activity. Accordingly, the question is not whether a correlation between an increase in mRNA and protein expression levels always exists, rather, if it is more likely than not to exist, and whether a person of ordinary skill in the pertinent art would recognize such a positive correlation.

Nowhere in the Lilley, Wildsmith or the King papers do the authors suggest that it is more likely than not that altered mRNA levels does not correlate with altered protein levels. In fact, the King reference discussed numerous advantages of the microarray technology, which offers tremendous advantages in the study of human diseases. For instance, on page 2287, the author states that “microarrays can be expected to prove extremely valuable as tools for the study of the generic basis of complex diseases. The ability to measure expression profiles across entire genomes provides a level of information not previously attainable..... Microarrays make it possible to investigate differential gene expression in normal vs. diseased tissue, in treated vs. non-treated tissue, and in different stages during the natural course of the disease, all on a genomic scale. Gene expression profiles may help to unlock the molecular basis of phenotype, response to treatment, and heterogeneity of disease.....” Therefore, if anything, the King reference supports the use of the microarray in the diagnosis of human diseases, which silently assumes that, most probably, increases in mRNA levels correlate well with increases in protein levels which in turn impacts disease.

Similarly, the Wildsmith paper discusses examples of a number of successes of microarray applications in the detection of human diseases (see Page 284). For instance, the author points out that “one area of rapid progress using microarray technology is the increased understanding of cancer. Molecular pathologies are subgrouping cancers of tissues such as blood, skin, and breast, based on differential gene expression patterns. For example, within a small group of breast cancer tissue samples, Perou *et al.* distinguished two broad subgroups representing those expressing or alternatively lacking expression of the oestrogen receptor-  $\gamma$ -gene. The work was not conclusive, but never has progress in this field been so rapid when compared with the previous methods of gene amplification. Another example of the impact of this technology is in the identification of two biomarkers for prostate cancer, namely hepsin and PIM1 (Dhanasekaran *et al.*, 2001). Microarray technology has also accelerated the understanding of the molecular events surrounding pulmonary fibrosis. Specially, two distinct clusters of genes associated with inflammation and fibrosis have been identified in a disease where, for years, the pathogenesis and treatment have remained unknown (Katsuma *et al.*, 2001).”



Therefore, contrary to the Examiner's position regarding the Lilley *et al.*, Lee *et al.*, Wildsmith *et al.*, and King *et al.* references, collectively, the references show that the art indicates that, generally, if a mRNA is overexpressed in cancer, it is more likely than not that the encoded protein will also be expressed at an elevated level.

**Madoz Gurpide *et al.* and Feroze-Merzoug *et al.***

Again, the Examiner cites Madoz-Gurpide *et al.* and Feroze-Merzoug *et al.* to show that one cannot accurately predict protein levels based on mRNA levels (made of record by the Examiner in the Final Office Action mailed October 17, 2006).

Appellants respectfully disagree. Madoz-Gurpide *et al.* explains that mRNA expression alone does not provide information regarding the **"activation state, post-translational modification or localization of corresponding proteins"** (emphasis added; page 168, col. 1). That is, Madoz-Gurpide *et al.* explain that mechanisms are not apparent from mRNA expression alone. Madoz Gurpide *et al.* further state that, it is "unclear" how **well** the reported RNA levels correlate with protein levels. In support of this assertion, the authors cite only a single reference, namely, the Chen *et al.*, which was discussed above and in detail in the Appeal Brief filed April 17, 2006. Madoz Gurpide *et al.* also acknowledge that the DNA microarray studies, such as those carried out by Beer *et al.* (specifically cited by the authors at page 52), **"justify the use of this technology for uncovering patterns of gene expression that are clinically informative"** (emphasis added; page 53). Thus, while Madoz-Gurpide *et al.* note that it is "more difficult to develop an **understanding of disease** at a mechanistic level using DNA microarrays," (emphasis added; page 53), Appellants respectfully point out that that "understanding of a disease at the mechanistic level" is not relevant to Appellants' assertions of utility, as discussed above. Accordingly, a *prima facie* case cannot be made based on the teachings within the Madoz-Gurpide *et al.* reference.

Similarly, Feroze-Merzoug *et al.* appear to mainly focus on "accurately predicting" the precise levels of protein expression, which is not required for utility as a cancer diagnostic, as discussed above. Moreover, the teachings of Feroze-Merzoug *et al.* are directed specifically to androgen regulated genes, which clearly involve different biological processes than those involved in tumor development. Feroze-Merzoug *et al.* are not directed to genes in general,

either, and hence this reference cannot be relied upon to establish a *prima facie* case for lack of utility.

**Nagaraja et al., Waghray et al., Sagynaliev et al.**

Again, new references Nagaraja *et al.*, Waghray *et al.*, Sagynaliev *et al.* all utilize and analyze the microarray assay. Any rejection directed to the limitations of the microarray assay or to the universal control are moot and are not addressed here. Only relevance of microarray array results to correlate mRNA to protein levels is discussed.

The Examiner asserts that “[c]omprehensive studies where significantly large numbers of transcripts and proteins were examined report that increases in mRNA and protein samples are not correlated.” (Final Office Action of October 17, 2006). The Examiner cites Nagaraja *et al.* as allegedly teaching that “the proteomic profiles indicated altered abundance of fewer proteins as compared to transcript profiles.” (Final Office Action of October 17, 2006).

Appellants respectfully submit that the fact that many more transcripts than proteins were found to be differentially expressed does not mean that most mRNA changes did not result in correlating protein changes, but merely reflects the fact that expression levels were only measured at all for many fewer proteins than transcripts. In particular, the total number of proteins whose expression levels could be visualized on silver-stained gels was only about 300 (page 2332, col. 1), as compared to the approximately 14,500 genes on the microarray chips for which mRNA levels were measured (page 2336, col. 1). Since the expression levels of so many fewer proteins than transcripts were measured, it is hardly surprising that a smaller absolute number of proteins than mRNAs were found to be overexpressed, because the protein products of most of the overexpressed mRNAs would not have been among the small number of proteins identified on the gels.

The Examiner next cites Waghray *et al.*, to the effect that “for most of the proteins identified, there was no appreciable concordant change at the RNA level,” and that “[t]he change in intensity for most of the affected proteins identified could not be predicted based on the level of the corresponding RNA.” (Final Office Action of October 17, 2006). Appellants reiterate that they need only show that there is a correlation between mRNA and protein levels, such that mRNA overexpression generally predict protein overexpression. A showing that mRNA levels can be used to “accurately predict” the precise levels of protein expression is not required.

Appellants also emphasize that Appellants are asserting that a measurable change in mRNA level generally leads to a corresponding change in the level of protein expression, not that changes in protein level can be used to predict changes in mRNA level. Waghray et al. did not take genes which showed significant mRNA changes and check the corresponding protein levels. Instead, the authors looked at a small and unrepresentative number of proteins, and checked the corresponding mRNA levels. Waghray et al. acknowledge that only “[a] relatively small set of genes could be analyzed at the protein level, largely due to the limited sensitivity of 2-D PAGE” (page 1337, col. 1). In particular, while the authors examined the expression levels of 16,570 genes (page 1329, col. 2), they were able to measure the expression levels of only 1031 proteins (page 1333, col. 2). Waghray et al. does not teach that changes in mRNA expression were not correlated with changes in expression of the corresponding protein. All Waghray et al. state is that “for most of the proteins identified, there was no appreciable concordant change at the mRNA level” (page 1337, col. 2). This statement is not relevant to Appellants’ assertion of utility, since Appellants are not asserting that changes in mRNA levels are the only cause of changes in protein levels. Waghray et al. do not contradict Appellants’ assertion that changes in mRNA expression, in general, correspond to changes in expression of the corresponding protein.

Lastly, the Examiner cites Sagynaliev et al., as allegedly teaching that “it is also difficult to reproduce transcriptomics results with proteomics tools.” In particular, the Examiner notes that according to Sagynaliev et al., of 982 genes found to be differentially expressed in human CRC, only 177 (18%) have been confirmed using proteomics technologies. (Final Office Action of October 17, 2006).

The Sagynaliev et al. reference, titled “Web-based data warehouse on gene expression in human colorectal cancer” (emphasis added), drew conclusions based upon a literature survey of gene expression data published in human CRC, and not from experimental data. While a literature survey can be a useful tool to assist researchers, the results may greatly over-represent or under-represent certain genes, and thus the conclusions may not be generally applicable. In particular, Appellants note that, as evidenced by Lian et al., Nagaraja et al., and Waghray et al., discussed above, the number of mRNAs examined in transcriptomics studies is typically much larger than the number of proteins examined in corresponding proteomics studies, due to the difficulties in detecting and

resolving more than a small minority of all expressed proteins on 2D gels. Thus the fact that only 18% of all genes found to be differentially expressed in human CRC have been confirmed using proteomics technologies does not mean that the corresponding proteins are not also differentially expressed, but is most likely due to the fact that the corresponding proteins were not identified on 2D gels, and thus their expression levels remain unknown.

The authors of Sagynaliev *et al.* acknowledge the many technical problems in finding proteomic data for CRC that can be matched to transcriptomic data to see if the two correlate. The authors state that “results have been obtained using heterogeneous samples in particular cell lines, whole tissue biopsies, and epithelial cells purified from surgical specimens.” However, “Results obtained in cell lines do not allow accurate comparison between normal and cancer cells, and the presence/absence of proteins of interest has to be confirmed in biopsies.” (Page 3072, left column.) In particular, the authors specifically note that “only a single study [1] provided differential display protein expression data obtained in the human patient, using whole tissue biopsy.” (Page 3068, left column, second paragraph; *see also*, Table 2.)

Appellants further note that Table 2 shows that 6 out of 8 published proteomics studies were done using 2-D PAGE. However, the authors state that “2-D PAGE or 2-D DIGE have well-known technological limitations ... even under well-defined experimental conditions, 2-D PAGE parallel analysis of paired CRC samples is hampered by a significant variability.” (Page 3077, left column, third paragraph.) Therefore, Appellants respectfully submit that it is well known in the art that there are problems associated with selecting only those proteins detectable by 2D gels.

Accordingly, the Examiner cannot rely on the teachings of Nagaraja *et al.*, Waghray *et al.*, Sagynaliev *et al.* to establish a *prima facie* showing of lack of utility.

#### Li *et al.*

The Examiner cites new reference Li *et al.* as teaching that “68.8% of the genes showing over-representation in the genome did not show elevated transcript levels.” (Final Office Action of October 17, 2006).

Appellants respectfully point out that Li *et al.* acknowledge that their results differed from those obtained by Hyman *et al.* and Pollack *et al.* (of record), who found a substantially higher level of correlation between gene amplification and increased gene expression. The authors note that “[t]his discordance may reflect methodologic differences between studies or

biological differences between breast cancer and lung adenocarcinoma” (page 2629, col. 1). For instance, as explained in the Supplemental Information accompanying the Li article, genes were considered to be amplified if they had a copy number ratio of at least 1.40. In the case of PRO232, as discussed in previously filed responses and in the Goddard Declaration (of record), an appropriate threshold for considering gene amplification to be significant is a copy number of at least 2.0 (which is a higher threshold). The PRO232 gene showed significant amplification of **2.056-fold to 5.28-fold, in five lung tumors** or **2.00-fold to 5.32-fold in seven colon tumors**, and thus fully meets this standard. It is not surprising that, in the Li *et al.* reference, by using a lower threshold of 1.4 for considering gene amplification, a higher number of genes not showing corresponding increases in mRNA expression were found. Nevertheless, the results of Li *et al.* do not conclusively disprove that a gene with a substantially higher level of gene amplification, such as PRO232, would be expected to show a corresponding increase in transcript expression.

Therefore, the Patent Office has failed to meet its initial burden of proof that Appellants’ claims of utility are not substantial or credible. The arguments presented by the Examiner based on references Pennica *et al.*, Konopka *et al.*, Haynes *et al.*, Hu *et al.*, Lian *et al.*, Fessler *et al.*, Gygi *et al.*, Lilley *et al.*, Lee *et al.*, King *et al.*, Wildsmith *et al.*, Nagaraja *et al.*, Sagynaliev *et al.*, Waghray *et al.*, Madoz-Gurpide *et al.*, Feroze-Merzoug *et al.*, Bustin *et al.*, Saito-Hisaminato *et al.*, do not provide sufficient reasons to doubt the statements by Appellants that PRO232 has utility as a diagnostic marker for lung or colon cancer. Appellants once again remind the Examiner that only after the Examiner has made a proper prima facie showing of lack of utility, does the burden of rebuttal shift to the Appellant. Based on the above discussions, such a showing has not been made. Accordingly, the instant rejection should be withdrawn for the Examiner’s lack of establishment of a *prima facie* showing.

Moreover, Appellants acknowledge that, in certain instances, DNA/mRNA and protein levels do not correlate. In fact, Appellants have included several such references directed towards a single gene or genes that lack a correlation in their IDS filed July 27, 2006. The references discussed in the Preliminary Amendment filed July 27, 2006, and the arguments therein are hereby incorporated by reference for brevity. The IDS included references that studied single genes or gene families, multiple or large families of genes, and included studies that a wide variety of techniques including gene amplification and microarray. Regardless of the

techniques employed, by and large, increased genes/transcripts levels mostly correlated with increased protein levels, even if accurate predictions of proteins could not be made. As discussed throughout prosecution, the law does not require the existence of a “necessary” correlation between DNA/mRNA and protein levels, or that protein levels be “accurately predicted”. In fact, authors in several of the cited references (cited both, by the Examiner, and by Appellants) themselves acknowledge that there is a general correlation between protein expression and transcript levels and DNA levels, which meets the “more likely than not standard”.

In summary, Appellants maintain that even though there are certain instances where a correlation, between DNA/mRNA and protein levels do not exist, in most cases, there is generally good correlation between them, and this was collectively demonstrated in the more than 100 references submitted by the Appellants in the IDS filed July 27, 2006.

**D. The Gene Amplification Data Establishes Credible, Substantial and Specific Patentable Utility for the PRO232 Polypeptide**

On the other hand, as discussed throughout prosecution, Appellants submit that Example 170 of the specification further discloses that, “(a)mplification is associated with overexpression of the gene product, indicating that the polypeptides are useful targets for therapeutic intervention in certain cancers such as lung, colon, breast and other cancers and diagnostic determination of the presence of those cancers” (Emphasis added). Appellants have also submitted ample evidence to show that, in general, if a gene is amplified in cancer, it is “more likely than not” that the encoded protein will also be expressed at an elevated level.

Besides the reference, the Declaration by Dr. Paul Polakis (Polakis I - made of record in Appellants' Response filed August 9, 2004), principal investigator of the Tumor Antigen Project of Genentech, Inc., the assignee of the present application, explains that in the course of Dr. Polakis' research using microarray analysis, he and his co-workers identified approximately 200 gene transcripts that are present in human tumor cells at significantly higher levels than in corresponding normal human cells. Appellants submit that Dr. Polakis' Declaration was presented to support the position that there is a correlation between mRNA levels and polypeptide levels. The second Declaration by Dr. Polakis (Polakis II) presented evidentiary data in Exhibit B. Exhibit B of the Declaration identified 28 gene transcripts out of 31 gene

transcripts (*i.e.*, greater than 90%) that showed good correlation between tumor mRNA and tumor protein levels. As Dr. Polakis' Declaration (Polakis II - made of record in Appellants' Response filed July 27, 2006) says "[a]s such, in the cases where we have been able to quantitatively measure both (i) mRNA and (ii) protein levels in both (i) tumor tissue and (ii) normal tissue, we have observed that in the vast majority of cases, there is a very strong correlation between increases in mRNA expression and increases in the level of protein encoded by that mRNA." Accordingly, Dr. Polakis has provided the facts to enable the Examiner to draw independent conclusions regarding protein data. Appellants further emphasize that the opinions expressed in the Polakis Declaration, including in the above quoted statement, are all based on factual findings. For instance, antibodies binding to about 30 of these tumor antigens were prepared and mRNA and protein levels were compared. In approximately 80% of the cases, the researchers found that increases in the level of a particular mRNA correlated with changes in the level of protein expressed from that mRNA when human tumor cells are compared with their corresponding normal cells. Therefore, Dr. Polakis' research, which is referenced in his Declaration, shows that, in general, there is a correlation between increased mRNA and polypeptide levels. Hence, one of skill in the art would reasonably expect that, based on the gene amplification data of the PRO232 gene, the PRO232 polypeptide is concomitantly overexpressed in lung or colon tumors studied as well.

Appellants further note that the sale of gene expression chips to measure mRNA levels is a highly successful business, with a company such as Affymetrix recording 168.3 million dollars in sales of their GeneChip® arrays in 2004. Clearly, the research community believes that the information obtained from these chips is useful (*i.e.*, that it is more likely than not that the results are informative of protein levels).

Thus, based on the asserted utility for PRO232 in the diagnosis of selected lung or colon tumors, the reduction to practice of the instantly claimed protein sequence of SEQ ID NO: 18 in the present application (also see pages 3- 4 and Example 4, page 149-150), the step-by-step preparation of chimeric PRO polypeptides, including those wherein the heterologous polypeptide is an epitope tag or an Fc region of an immunoglobulin (page 74, lines 23 to page 75, line 5), the description of the expression of PRO polypeptides in various host cells, including *E. coli*, mammalian cells, yeast and Baculovirus-infected insect cells at least in Examples 53-56, pages

192-199, the disclosure of the step-by-step protocol for the preparation, isolation and detection of monoclonal, polyclonal and other types of antibodies against the PRO232 protein in the specification (monoclonal and polyclonal antibodies at page 139, line 32, to page 141, line 13; humanized antibodies at page 141, line 15, to page 142, line 16; antibody fragments at page 143, line 8 onwards; labeled antibodies at pages 144-145, line 16 onwards and page 146, line 33 to page 147, line 6) and the disclosure of the gene amplification assay in Example 92, the skilled artisan would know exactly how to make and use the claimed polypeptides for the diagnosis of lung or colon cancers. Appellants submit that based on the detailed information presented in the specification and the advanced state of the art in oncology, the skilled artisan would have found such testing routine and not 'undue.'

Therefore, since the instantly claimed invention is supported by either a credible, specific and substantial asserted utility or a well-established utility, one skilled in the art would know "how to make and use" the claimed invention without undue experimentation, Appellants respectfully request reconsideration and reversal of the determination of priority for Claims 44-46 and 49-51.

**ISSUE 2. Claims 44-46 and 49-51 should be entitled to the priority date of U.S. Provisional Application 60/059121, filed September 17, 1997**

The instant application has not been granted the earlier priority date of U.S. Provisional Application 60/059121, filed September 17, 1997 on the grounds that the prior 60/059121 application fails to provide a utility and lacks an enabling disclosure for the claimed invention under 35 U.S.C. §§101/112, first paragraph."

Appellants disagree and submit that, for the same reasons discussed above under Issue 1, U.S. Provisional Application 60/059121 also satisfies the utility requirements. Therefore, Appellants should be entitled to the priority date of **September 17, 1997**.

**ISSUE 3. Claims 44-46 and 49-51 are not anticipated by Rosenthal *et al.*, DE19818619-A1 (dated 10/1999)**

Claims 44-46 and 49-51 remain rejected under 35 U.S.C. §102(b) as being anticipated by Rosenthal *et al.*, DE19818619-A1 (dated 10/1999).



For the reasons discussed above under Issue 2, Appellants maintain that they are entitled to an effective filing date of September 17, 1997 based on a properly claimed priority to International application PCT/US98/18824. Therefore, Rosenthal *et al.* is not prior art and does not anticipate the instant claims. Accordingly, this rejection under 35 U.S.C. §102(b) should be withdrawn.


### **CONCLUSION**

For the reasons given above, Appellants submit that present specification and the specification of U.S. Provisional Application 60/059121 dated September 17, 1997 clearly describes and provides at least one patentable utility for the instantly claimed invention. Moreover, it is respectfully submitted that the present specification clearly teaches "how to use" the presently claimed polypeptide based upon this disclosed patentable utility. Accordingly, Rosenthal *et al.*, DE19818619-A1 is not prior art. As such, Appellants respectfully request reconsideration and reversal of the outstanding rejection of claims 44-46 and 49-51.

The Commissioner is authorized to charge any fees which may be required, including extension fees, or credit any overpayment to Deposit Account No. **08-1641** (referencing Attorney's Docket No. **39780-1618 P2C18**).

Respectfully submitted,

Date: July 23, 2007

  
Ginger R. Dreger  
Reg. No. 33,055

**HELLER EHRMAN LLP**  
275 Middlefield Road  
Menlo Park, California 94025-3506  
Telephone: (650) 324-7000  
Facsimile: (650) 324-0638

## VIII. CLAIMS APPENDIX

### Claims on Appeal

44. An isolated polypeptide comprising:

- (a) the amino acid sequence of the polypeptide of SEQ ID NO:18;
- (b) the amino acid sequence of the polypeptide of SEQ ID NO:18, lacking its

associated signal peptide; or

- (c) the amino acid sequence of the polypeptide encoded by the full-length coding sequence of the cDNA deposited under ATCC accession number 209250;

wherein, the nucleic acid encoding said polypeptide is amplified in lung or colon tumors.

45. The isolated polypeptide of Claim 44 comprising the amino acid sequence of the polypeptide of SEQ ID NO:18.

46. The isolated polypeptide of Claim 44 comprising the amino acid sequence of the polypeptide of SEQ ID NO:18, lacking its associated signal peptide.

49. The isolated polypeptide of Claim 44 comprising the amino acid sequence of the polypeptide encoded by the full-length coding sequence of the cDNA deposited under ATCC accession number 209250.

50. A chimeric polypeptide comprising a polypeptide according to Claim 44 fused to a heterologous polypeptide.

51. The chimeric polypeptide of Claim 50, wherein said heterologous polypeptide is an epitope tag or an Fc region of an immunoglobulin.

## IX. EVIDENCE APPENDIX

1. Declaration of Audrey Goddard, Ph.D. under 35 C.F.R. §1.132, with attached Exhibits

A-G:

- A. Curriculum Vitae of Audrey D. Goddard, Ph.D.
  - B. Higuchi, R. *et al.*, "Simultaneous amplification and detection of specific DNA sequences," *Biotechnology* 10:413-417 (1992).
  - C. Livak, K.J., *et al.*, "Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization," *PCR Methods Appl.* 4:357-362 (1995).
  - D. Heid, C.A. *et al.*, "Real time quantitative PCR," *Genome Res.* 6:986-994 (1996).
  - E. Pennica, D. *et al.*, "WISP genes are members of the connective tissue growth factor family that are up-regulated in Wnt-1-transformed cells and aberrantly expressed in human colon tumors," *Proc. Natl. Acad. Sci. USA* 95:14717-14722 (1998).
  - F. Pitti, R.M. *et al.*, "Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer," *Nature* 396:699-703 (1998).
  - G. Bieche, I. *et al.*, "Novel approach to quantitative polymerase chain reaction using real-time detection: Application to the detection of gene amplification in breast cancer," *Int. J. Cancer* 78:661-666 (1998).
2. Declaration of Avi Ashkenazi, Ph.D. under 35 C.F.R. §1.132, with attached Exhibit A (Curriculum Vitae).
3. Declaration of Paul Polakis, Ph.D. under 35 C.F.R. §1.132 (Polakis I).
4. Declaration of Paul Polakis, Ph.D. under 35 C.F.R. §1.132 (Polakis II).
5. Haynes *et al.*, "Proteome analysis: Biological assay or data archive?" *Electrophoresis* 19:1862-1871 (1996).
6. Rosenthal *et al.*, DE19818619-A1 (dated October 28, 1999).
7. Hyman, E., *et al.*, "Impact of DNA Amplification on Gene Expression Patterns in Breast Cancer," *Cancer Research* 62:6240-6245 (2002).
8. Pollack, J.R., *et al.*, "Microarray Analysis Reveals a Major Direct Role of DNA Copy Number Alteration in the Transcriptional Program of Human Breast Tumors," *Proc. Natl. Acad. Sci. USA* 99:12963-12968 (2002).
9. Hanna *et al.*, "HER-2/neu Breast Cancer Predictive Testing," Pathology Associates Medical Laboratories (1999).
10. Konopka *et al.*, "Variable Expression of the Translocated c-abl oncogene in Philadelphia-chromosome-positive B-lymphoid cell lines from chronic myelogenous leukemia patients" *Proc. Natl. Acad. Sci. USA* 83: 4049-52, (1986).

11. Hu *et al.*, "Analysis of genomic and proteomic data using advanced literature mining," *J. Proteome Res.* 2: 405-412 (2003).
12. Lian *et al.*, "Genomic and proteomic analysis of the myeloid differentiation program," *Blood* 98: 513-524 (2001).
13. Fessler *et al.*, "A genomic and proteomic analysis of activation of the human neutrophil by lipopolysaccharide and its mediation by p38 mitogen-activated protein kinase," *J. Biol. Chem.* 277: 31291-31302 (2002).
14. Abe, N., *et al.*, *Br J Cancer* - 89(11):2104-9 (2003).
15. Alberts, B., *et al.*, *Molecular Biology of the Cell* (3rd ed. 1994) Cell 3rd at 453 Figure 9-2 of Cell 3rd Cell 3rd at 403.
16. Alberts, B., *et al.*, *Molecular Biology of the Cell* (4rd ed.) In Cell 4th, Figure 6-3 on page 302 Figure 6-90 on page 364 of Cell 4th Cell 4th at 364 Cell 4th at 379.
17. Ando, M., *et al.*, *Br J Haematol.*, - 130(6):860-8 (2005).
18. Aust, G., *et al.*, *Thyroid* - 7(5):713-24 (1997).
19. Barnes, V.L., *et al.*, *J Histochem Cytochem.* - 47(6):787-98 (1999).
20. Bea, S., *et al.*, *Cancer Res.* - 61(6):2409-12 (2001).
21. Beer, *et al.*, *Nature Medicine* - 8(8):816-824 (2002).
22. Blaschke, V., *et al.*, *J Immunol Methods.* - 246(1-2):79-90 (2000).
23. Buckley, A.R., *et al.*, *Apoptosis.* - 2(6):518-28 (1997).
24. Caberlotto, L. *et al.*, *Neurosci Lett.* - 265(3):191-4 (1999).
25. Caberlotto, L., *et al.*, *Eur J Neurosci.* - 17(9):1736-46 (2003).
26. Choi, D., *et al.*, *J Soc Gynecol Investig.* - 9(1):41-6 (2002).
27. Couvelard, A., *et al.*, *Virchows Arch.* - 445(3):279-84 (2004).
28. Dagenais, A., *et al.*, *Am J. Physiol Lung Cell Mol Physiol.* - 286(2):L301-11 (2004).
29. De Boer, C.J., *et al.*, *Ann Oncol.* - 7(3):251-6 (1996).
30. Debieve, F., *et al.*, *Mol Hum Reprod.* - 6(8):743-9 (2000).
31. Dong, Z., *et al.*, *Invest Ophthalmol Vis Sci* - 42(13):3223- (2001).
32. Duchrow, M., *et al.*, *Cancer Invest.* - 19(6):588-96 (2001).

33. Dyer, J., *et al.*, Equine Vet J. - 34(4):349-58 (2002).
34. Egwuagu, C.E., *et al.*, J Immunol. - 168(7):3181-7 (2002).
35. El-Ghrably, I.A., *et al.*, Br J Ophthalmol. - 85(4):461-70 (2001).
36. Eleore, L., *et al.*, Neuroscience - 136(1):147-60 (2005).
37. Forsberg, H., *et al.*, Free Radic Res. - 24(6):451-9 (1996).
38. Freyschuss, B., *et al.*, Endocrinology - 133(4):1548-54 (1993).
39. Fu, K., *et al.*, Blood - 106(13):4315-21 (2005).
40. Fuchs, A.R., *et al.*, Biol Reprod. - 54(3):700-8 (1996).
41. Furuta, J., *et al.*, Melanoma Res. - 15(1):15-20 (2005).
42. Futcher, B., *et al.*, Mol Cell Biol., - 19(11):7357-68 (1999).
43. George, J., *et al.*, Biochem Pharmacol. - 49(7):873-81 (1995).
44. Giroux, M., *et al.*, J Immunol. - 165(7):3985-91 (2000).
45. Gnatenko, D.V., *et al.*, Blood - 101(6):2285-93 (2003).
46. Godbout, R., *et al.*, J Biol Chem, - 273(33):21161-8 (1998).
47. Goldenberg, R.C., *et al.*, J Endocrinol. - 177(2):327-35 (2003).
48. Golebiowski, F., *et al.*, Acta Biochim Pol. - 50(4):1065-73 (2003).
49. Greenbaum, D., *et al.*, BioinformaticsI - 18(4):585-496 (2002).
50. Grem, J.L., *et al.*, Clin Cancer Res. - 7(4):999-1009 (2001).
51. Grenback, E., *et al.*, Regul Pept, - 117(2):127-39 (2004).
52. Gromova, I., *et al.*, Int J. Oncol. - 13(2):379-83 (1998).
53. Guo, Y., *et al.*, Zhinghua Jie He He Hu Xi Za Zhi - 25(6):337-40 (2002).
54. Habu, Y., *et al.*, Biochem Pharmacol - 69(6):993-9 (2005).
55. Hahn, M.E., *et al.*, Toxicol Appl Pharmacol. - 127(2):187-98 (1994).
56. Hähnel, E., *et al.*, Breast Cancer Research and Treatment - 24:71-74 (1992).

57. Hamilton, L.M., *et al.*, Clin Exp Allergy. - 33(2):233-40 (2003).
58. Hassett, C., *et al.*, Arch Biochem Biophys. - 337(2):275-83 (1997).
59. Hirsch, F.R., *et al.*, Clinical Cancer Research - 7:5-22 (2001).
60. Holten-Andersen, M.N., *et al.*, Int J Cance, - 113(2):198-206 (2005).
61. Huang, Y.H., *et al.*, J Med Virol. - 70(4):610-6 (2003).
62. Huettner, P.C., *et al.*, Mod Pathol. - 5(3):250-6 (1992).
63. Hui, P., *et al.*, Leuk Lymphoma. 44(8):1385-94 (2003).
64. Husain, I., *et al.*, Cancer Res. - 54(2):539-46 (1994).
65. Ihmann, T., *et al.*, J Cancer Res Clin Oncol. - 130(12):749-56 (2004).
66. Ikegami, T., *et al.*, Am J Physiol Cell Physiol. - 281(4):C1396-402 (2001).
67. Jacquemin, E., *et al.*, Int J Biol. - 37(3):417-23 (1993).
68. Jaime, M., *et al.*, Hepatology - 35(5):1063-71 (2002).
69. Janssens, N., *et al.*, Tumour Biol. - 25(4):161-71 (2004).
70. Jungbluth, A.A., *et al.*, Int J. Cancer - 92(6):856-60 (2001).
71. Kalabis, G.M., *et al.*, Biol Reprod. - 73(4):591-7 (2005).
72. Kammaori, M., *et al.*, Int J Oncol. - 27(5):1257-63 (2005).
73. Khal, J., *et al.*, Br J Cancer - 93(7):774-80 (2005).
74. Khal, J., *et al.*, Int J. Biochem Cell Biol. - 37(10):2196-206 (2005).
75. Kogo, H., *et al.*, J Biol Chem. - 279(24):25574-81 (2004).
76. Kommoss, F., *et al.*, Acta Obstet Gynecol Scand Suppl. - 155:19-24 (1992).
77. Kumar, U., *et al.*, Breast Cancer Res. Treat. - 92(2):175-86 (2005).
78. Kuo, C.C., *et al.*, Proteomics - 5(4):894-906 (2005).
79. Landmark, B.F., *et al.*, J Reprod Fertil. - 99(2):323-34 (1993).
80. Lassmann, S., *et al.*, J Pathol. - 198(2):198-206 (2002).

81. Legrand, O., *et al.*, Br J, Haematol. - 94(1):23-33 (1996).
82. Lemstrom, K.B., *et al.*, Circulation - 105(21):2524-30 (2002).
83. Lewin, B., Genes VI Genes VI at 847-848 (1997).
84. Li, Y., *et al.*, Invest Ophthalmol Vis Sci. - 44(3):1299-304 (2003).
85. Li, Z.B., *et al.*, Biochem Biophys Res Commun. - 327(4):1163-9 (2005).
86. Lindberg, P., *et al.*, Arch Oral Biol. - 46(1):23-31 (2001).
87. Macabeo-Ong, M., *et al.*, 15(9):979-87 (2002).
88. Maruyama, H., *et al.*, Am J Pathol. - 155(3):815-22 (1999).
89. Meehan, T.P., *et al.*, J Mol Endocrinol. - 32(1):247-55 (2004).
90. Mendoza-Rodriguez, C.A., *et al.*, Mol Reprod Dev. 64(4):379-88 (2003).
91. Meoni, P., *et al.*, Brain Res Mol Res. - 54(1):13-23 (1998).
92. Meric, F., *et al.*, Molecular Cancer Therapeutics - 1:971-979 (2002).
93. Mezzano, S.A., *et al.*, Kiney Int. - 57(1):147-58 (2000).
94. Mingrone, G., *et al.*, Obes Res. - 11(5):632-40 (2003).
95. Miralles, C.P., *et al.*, Brain Res Mol Res. - 24(1-4):129-39 (1994).
96. Mizrachi, D., *et al.*, Biol Reprod. - 61(3):776-84 (1999).
97. Monaghan, P., *et al.*, J Gen Virol. - 86(Pt 10):2769-80 (2005).
98. Montuori, N., *et al.*, Int J Cancer - 105(3):353-60 (2003).
99. Munaut, C., *et al.*, Int J Cancer - 106(6):848-55 (2003).
100. Nie, Y., *et al.*, Carcinogenesis - 22(10):1615-23 (2001).
101. Nuciforo, P.G., *et al.*, Hum Pathol. - 34(7):639-45 (2003).
102. Oberringer, M., *et al.*, Biochem Biophys Res Commun. - 214(3):1009-14 (1995).
103. Ornftoft, T.F., *et al.* Molecular & Cellular Proteomics – 1:37-45 (2002).
104. Pachmann, K., *et al.*, Br J. Haematol - 112(3):749-59 (2001).

105. Pairon, J.C., *et al.*, Am J Respir Cell Mol Biol. - 11(4):386-96 (1994).
106. Papotti, M., *et al.*, Diagn Mol Pathol. - 9(1):47-57 (2000).
107. Papotti, M., *et al.*, Virchows Arch. - 440(5):461-75 (2002).
108. Paredes, J., *et al.*, Clin Cancer Res. - 11(16):5869-77 (2005).
109. Politis, I., *et al.*, J Dairy Sci. - 75(6):1423-9 (1992).
110. Preesman, A.H., *et al.*, J Invest Dermatol. - 99(5):587-93 (1992).
111. Pullig, O., *et al.*, Osteoarthritis Cartilage - 10(4):253-63 (2002).
112. Rey, C., *et al.*, Biochem Pharmacol. - 60(11):1636-46 (2000).
113. Rudlowski, C., *et al.*, Am J. Clin Pathol. - 120(5):691-8 (2003).
114. Sasaki, T., *et al.*, Exp Cell Res. - 275(2):189-99 (2002).
115. Sedelies, K.A., *et al.*, J Biol Chem. - 279(25):26581-7 (2004).
116. Shen, Y., *et al.*, Blood - 104(9):2936-9 (2004).
117. Shinohara, Y., *et al.*, Biochim Biophys Acta - 1368(1):129-36 (1998).
118. Silvers, A.L., *et al.*, Photochem Photobiol. - 75(3):302-10 (2002).
119. Song, L., *et al.*, Am J. Physiol. - 267(4 Pt 2):F546-57 (1994).
120. Spaziani, E.P., *et al.*, J Interferon Cytokine Res. - 18(12):1039-44 (1998).
121. Spika, I., *et al.*, Skin Pharmacol Appl Skin Physiol. - 16(3):143-50 (2003).
122. Splinter, P.L., *et al.*, J Biol Chem - 278(8):6268-74 (2003).
123. Stearns, M.E., *et al.*, Cancer Res. - 53(4):878-83 (1993).
124. Stein, R., *et al.*, J Urol. - 164(3Pt 2):1026-30 (2000).
125. Strickland, I., *et al.*, J Invest Dermatol. - 108(5):763-8 (1997).
126. Strutz, F., *et al.*, Kidney Int. - 57(4):1521-38 (2000).
127. Takahashi, K., *et al.*, J Biol. Chem. - 278(47):46654-60 (2003).
128. Takimoto, Y., *et al.*, Circulation - 105(4):490-6 (2002).



129. Telek, G., *et al.*, J Surg Res. - 96(1):56-67 (2001).
130. Timchenko, L., *et al.*, Semin Cell Biol. - 6(1):13-9 (1995).
131. Torronen, R., *et al.*, Chem Biol. Interact. - 83(2):107-19 (1992).
132. Ullmannova, V., *et al.*, Leuk Res. - 27(12):1115-23 (2003).
133. Van Beers, E.H., *et al.*, J Histochem Cytochem. - 46(2):231-40 (1998).
134. Van Der Wilt, C.L., *et al.*, Eur J Cancer - 39(5):691-7 (2003).
135. Waldherr, R., *et al.*, Pediatr Nephrol. - 7(4):471-8 (1993).
136. Walmer, D.K., *et al.*, Cancer Res. - 55(5):1168-75 (1995).
137. Wang, J., *et al.*, Cancer Res. - 54(2):560-4 (1994).
138. Wang, J., *et al.*, Urol Res. - 28(5):308-15 (2000).
139. Wang, L.G., *et al.*, Cancer Res. - 57(4):714-9 (1997).
140. Weterman, M.A., *et al.*, Cancer Res. - 53(24):6061-6 (1993).
141. Williams, E.T., *et al.*, J Pharmacol Exp Ther. 311(2):728-35 (2004).
142. Wojtaszek, P.A., *et al.*, Oncogene - 8(3):755-60 (1993).
143. Zhigang, Z., *et al.*, World Journal of Surgical Oncology - 2-13 (2004).
144. Zhong, W., *et al.*, Free Radic Biol Med. - 27(11-12):1334-45 (1999).
145. Xi, L., *et al.*, Zhonghua Fu Chan Ke Za Zhi - 40(6):407-10 (2005).
146. Bustin *et al.*, (2002) Trends in Mol. Med., 8(6): 269-272.
147. Lilley *et al.*, "Proteomics" Mol. Biol. in Cellular Path. (2003) England: John Wiley & Sons, p 351-352.
148. Lee *et al.*, (2000) Proc. Nat. Acad. Sci., USA, 97(18): 9834-39.
149. King *et al.*, 2001, J. Am. Med. Assoc. 286: 2280-2288.
150. Wildsmith *et al.*, "Gene Expression Analysis using Microarrays" Mol. Biol in Cellular Path. (2003) England: John Wiley & Sons, p 269-286.
151. Nagaraja *et al.*, 2006, Oncogene 25: 2328-2338.

152. Waghray *et al.*, 2001, Proteomics, 1: 1327-1338.
153. Sagynaliev *et al.*, 2005, Proteomics, 5: 3066-3076.
154. Feroze-Merzoug *et al.*, (2001), Cancer and Metastasis Rev., 20:165-171.
155. Madoz-Gurpide *et al.*, 2003, Adv. Exp. Med. Biol., 532:51-58.
156. Saito-Hisaminato *et al.*, (2002), DNA Research, 9:35-45.
157. Gygi *et al.*, (1999), Mol. and Cell Biol., 19: 1720-30.
158. Li *et al.*, 2006, Oncogene, 25: 2628-2635.

Item 1 was submitted with Appellants' Response filed July 25, 2005, and were considered by the Examiner as indicated in the second Final Office action mailed October 21, 2005.

Item 2 was submitted with Appellants' Response filed December 10, 2003, and was considered by the Examiner as indicated in the Office action mailed April 8, 2004.

Items 3, 103 and 7-9, were submitted with Appellants' Response filed August 9, 2004, and were considered by the Examiner as indicated in the Office action mailed December 6, 2004.

Item 4 was submitted with Appellants' Response filed July 27, 2006, and was considered by the Examiner as indicated in the Final Office Action mailed October 17, 2006.

Item 1 (E), i.e.: Pennica *et al.*, and Items 5-6 were made of record by the Examiner in the Office Action mailed September 29, 2003.

Items 10-11 were made of record by the Examiner in the Office Action mailed February 23, 2005.

Items 12-13 were made of record by the Examiner in the Final Office Action mailed October 21, 2005.

Items 14-145 were submitted with Appellants' Response filed July 27, 2006, and were considered by the Examiner as indicated in the Final Office Action mailed October 17, 2006.

Items 146-158 were made of record by the Examiner in the Final Office Action mailed October 17, 2006.

**X. RELATED PROCEEDINGS APPENDIX**

None- no decision rendered by a Court or the Board in any related proceedings identified above.



⑮ BUNDESREPUBLIK  
DEUTSCHLAND



DEUTSCHES  
PATENT- UND  
MARKENAMT

⑫ **Offenlegungsschrift**  
⑩ **DE 198 18 619 A 1**

⑳ Aktenzeichen: 198 18 619.3  
㉑ Anmeldetag: 21. 4. 98  
㉒ Offenlegungstag: 28. 10. 99

⑤① Int. Cl.<sup>6</sup>:  
**C 07 K 16/00**  
C 07 K 14/435  
A 61 K 38/17  
C 07 H 21/04  
C 12 N 15/11  
C 12 N 15/63  
C 12 N 1/21  
C 12 N 1/19  
C 12 N 5/10  
// (C12N 1/21, C12R  
1:19)G01N 33/68,  
33/15

DE 198 18 619 A 1

⑦① Anmelder:  
metaGen Gesellschaft für Genomforschung mbH,  
14195 Berlin, DE  
  
⑦④ Vertreter:  
Klose, W., Dipl.-Chem.Dr.rer.nat., Pat.-Ass., 13505  
Berlin

⑦② Erfinder:  
Rosenthal, André, Prof. Dr., 10115 Berlin, DE;  
Specht, Thomas, Dr., 12163 Berlin, DE; Hinzmann,  
Bernd, Dr., 13127 Berlin, DE; Schmitt, Armin, Dr.,  
14197 Berlin, DE; Pilarsky, Christian, Dr., 14532  
Stahnsdorf, DE; Dahl, Edgar, Dr., 14480 Potsdam,  
DE

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen

⑤④ Menschliche Nukleinsäuresequenzen aus Blase-Tumor

⑤⑦ Es werden menschliche Nukleinsäuresequenzen -  
mRNA, cDNA, genomische Sequenzen - aus Blasentu-  
morgewebe, die für Genprodukte oder Teile davon kodie-  
ren, und deren Verwendung beschrieben.  
Es werden weiterhin die über die Sequenzen erhältlichen  
Polypeptide und deren Verwendung beschrieben.

DE 198 18 619 A 1

Die Erfindung betrifft menschliche Nukleinsäuresequenzen aus Blasenmorgewebe, die für Genprodukte oder Teile davon kodieren, deren funktionale Gene, die mindestens ein biologisch aktives Polypeptid kodieren und deren Verwen-

5 dung.

Die Erfindung betrifft weiterhin die über die Sequenzen erhältlichen Polypeptide und deren Verwendung.

Eine der Hauptkrebstodesursachen ist der Blasenmorgewebe, für dessen Bekämpfung neue Therapien notwendig sind. Bisher verwendete Therapien, wie z. B. Chemotherapie, Hormontherapie oder chirurgische Entfernung des Tumorgewebes, führen häufig nicht zu einer vollständigen Heilung.

10 Das Phänomen Krebs geht häufig einher mit der Über- oder Unterexpression gewisser Gene in den entarteten Zellen, wobei noch unklar ist, ob diese veränderten Expressionsraten Ursache oder Folge der malignen Transformation sind. Die Identifikation solcher Gene wäre ein wesentlicher Schritt für die Entwicklung neuer Therapien gegen Krebs. Der spontanen Entstehung von Krebs geht häufig eine Vielzahl von Mutationen voraus. Diese können verschiedenste Auswirkungen auf das Expressionsmuster in dem betroffenen Gewebe haben, wie z. B. Unter- oder Überexpression, aber auch Expression verkürzter Gene. Mehrere solcher Veränderungen durch solche Mutationskaskaden können schließlich zu bös-

15 artigen Entartungen führen. Die Komplexität solcher Zusammenhänge erschwert die experimentelle Herangehensweise sehr.

Für die Suche nach Kandidatengen, d. h. Genen, die im Vergleich zum Tumorgewebe im normalen Gewebe stärker exprimiert werden, wird eine Datenbank verwendet, die aus sogenannten ESTs besteht. ESTs (Expressed Sequence Tags) sind Sequenzen von cDNAs, d. h. revers transkribierten mRNAs, den Molekülen also, die die Expression von Genen widerspiegeln. Die EST-Sequenzen werden für normale und entartete Gewebe ermittelt. Solche Datenbanken werden von verschiedenen Betreibern z. T. kommerziell angeboten. Die ESTs der LifeSeq-Datenbank, die hier verwendet wird, sind in der Regel zwischen 150 und 350 Nukleotide lang. Sie repräsentieren ein für ein bestimmtes Gen unverkennbares Muster, obwohl dieses Gen normalerweise sehr viel länger ist (> 2000 Nukleotide). Durch Vergleich der Expressionsmuster von normalen und Tumorgewebe können ESTs identifiziert werden, die für die Tumorentstehung und -proliferation wichtig sind. Es besteht jedoch folgendes Problem: Da durch unterschiedliche Konstruktionen der cDNA-Bibliotheken die gefundenen EST-Sequenzen zu unterschiedlichen Regionen eines unbekannten Gens gehören können, ergäbe sich in einem solchen Fall ein völlig falsches Verhältnis des Vorkommens dieser ESTs in dem jeweiligen Gewebe. Dieses würde erst bemerkt werden, wenn das vollständige Gen bekannt ist und somit die ESTs dem gleichen Gen zugeordnet werden können.

Es wurde nun gefunden, daß diese Fehlermöglichkeit verringert werden kann, wenn zuvor sämtliche ESTs aus dem jeweiligen Gewebetyp assembliert werden, bevor die Expressionsmuster miteinander verglichen werden. Es wurden also überlappende ESTs ein und desselben Gens zu längeren Sequenzen zusammengefaßt (s. Fig. 1, Fig. 2a und Fig. 3). Durch diese Verlängerung und damit Abdeckung eines wesentlich größeren Genbereichs in jeder der jeweiligen Banken sollte der oben beschriebene Fehler weitgehendst vermieden werden. Da es hierzu keine bestehenden Softwareprodukte gab, wurden Programme für das Assemblieren von genomischen Abschnitten verwendet, die abgewandelt eingesetzt und durch eigene Programme ergänzt wurden. Ein Flowchart der Assemblierungsprozedur ist in Fig. 2b1-2b4 dargestellt.

Es konnten nun die Nukleinsäure-Sequenzen Seq. ID No. 1 50 gefunden werden, die als Kandidatengene beim Blasenmorgewebe eine Rolle spielen.

40 Von besonderem Interesse sind die Nukleinsäure-Sequenzen Seq. ID Nos. 2-5, 7-13, 16, 18, 20, 23, 26-27, 31-32, 36, 45.

Die Erfindung betrifft somit Nukleinsäure-Sequenzen, die ein Genprodukt oder ein Teil davon kodieren, umfassend

- 45 a) eine Nukleinsäure-Sequenz, ausgewählt aus der Gruppe der Nukleinsäure-Sequenzen Seq. ID Nos. 2-5, 7-13, 16, 18, 20, 23, 26-27, 31-32, 36, 45.
- b) eine allelische Variation der unter a) genannten Nukleinsäure-Sequenzen
- oder
- c) eine Nukleinsäure-Sequenz, die komplementär zu den unter a) oder b) genannten Nukleinsäure-Sequenzen ist.

50 Die Erfindung betrifft weiterhin eine Nukleinsäure-Sequenz gemäß einer der Sequenzen Seq. ID Nos. 2-5, 7-13, 16, 18, 20, 23, 26-27, 31-32, 36, 45 oder eine komplementäre oder allelische Variante davon und die Nukleinsäure-Sequenzen davon, die eine 90%ige bis 95%ige Homologie zu einer humanen Nukleinsäure-Sequenz aufweisen.

Die Erfindung betrifft auch die Nukleinsäure-Sequenzen Seq. ID No. 1 bis Seq. ID No. 50, die im Blasenmorgewebe erhöht exprimiert sind.

55 Die Erfindung betrifft ferner Nukleinsäure-Sequenzen, umfassend einen Teil der oben genannten Nukleinsäure-Sequenzen, in solch einer ausreichenden Größe, daß sie mit den Sequenzen Seq. ID Nos. 1-50 hybridisieren.

Die erfindungsgemäßen Nukleinsäure-Sequenzen weisen im allgemeinen eine Länge von mindestens 50 bis 4500 bp, vorzugsweise eine Länge von mindestens 150 bis 4000 bp, insbesondere eine Länge von 450 bis 3500 bp auf.

60 Mit den erfindungsgemäßen Teilsequenzen Seq. ID Nos. 1-50 können gemäß gängiger Verfahrenspraxis auch Expressionskassetten konstruiert werden, wobei auf der Kassette mindestens eine der erfindungsgemäßen Nukleinsäure-Sequenzen zusammen mit mindestens einer dem Fachmann allgemein bekannten Kontroll- oder regulatorischen Sequenz, wie z. B. einem geeigneten Promotor, kombiniert wird. Die erfindungsgemäßen Sequenzen können in sense oder antisense Orientierung eingefügt sein.

In der Literatur sind eine große Anzahl von Expressionskassetten bzw. Vektoren und Promotoren bekannt, die verwendet werden können.

65 Unter Expressionskassetten bzw. Vektoren sind zu verstehen: 1. bakterielle, wie z. B., phagescript, pBs,  $\phi$ X174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene), pTrec99A, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia), 2. eukaryontische, wie z. B. pWLneo, pSV2cat, pOG44, pXT1, pSG (Stratagene), pSVK3, pBPV, pMSG,

pSVL (Pharmacia).

Unter Kontroll- oder regulatorischer Sequenz sind geeignete Promotoren zu verstehen. Hierbei sind zwei bevorzugte Vektoren der pKK232-8 und der PCM7 Vektor. Im einzelnen sind folgende Promotoren gemeint: lacI, lacZ, T3, T7, gpt, lambda P<sub>R</sub>, trc, CMV, HSV Thymidin-Kinase, SV40, LTRs aus Retrovirus und Maus Metallothionein-I.

Die auf der Expressionskassette befindlichen DNA-Sequenzen können ein Fusionsprotein kodieren, das ein bekanntes Protein und ein biologisch aktives Polypeptid-Fragment umfaßt.

Die Expressionskassetten sind ebenfalls Gegenstand der vorliegenden Erfindung.

Die erfindungsgemäßen Nukleinsäure-Fragmente können zur Herstellung von Vollängen-Genen verwendet werden. Die erhältlichen Gene sind ebenfalls Gegenstand der vorliegenden Erfindung.

Die Erfindung betrifft auch die Verwendung der erfindungsgemäßen Nukleinsäure-Sequenzen, sowie die aus der Verwendung erhältlichen Gen-Fragmente.

Die erfindungsgemäßen Nukleinsäure-Sequenzen können mit geeigneten Vektoren in Wirtszellen gebracht werden, in denen als heterologer Teil die auf den Nukleinsäure-Fragmenten enthaltene genetischen Information befindet, die expri-  
miert wird.

Die die Nukleinsäure-Fragmente enthaltenden Wirtszellen sind ebenfalls Gegenstand der vorliegenden Erfindung.

Geeignete Wirtszellen sind z. B. prokaryontische Zellsysteme wie E. coli oder eukaryontische Zellsysteme wie tieri-  
sche oder humane Zellen oder Hefen.

Die erfindungsgemäßen Nukleinsäure-Sequenzen können in sense oder antisense Form verwendet werden.

Die Herstellung der Polypeptide oder deren Fragment erfolgt durch Kultivierung der Wirtszellen gemäß gängiger Kul-  
tivierungsmethoden und anschließender Isolierung und Aufreinigung der Peptide bzw. Fragmente, ebenfalls mittels gän-  
giger Verfahren. Die Erfindung betrifft ferner Nukleinsäure-Sequenzen, die mindestens eine Teilsequenz eines biolo-  
gisch aktiven Polypeptids kodieren.

Ferner betrifft die vorliegende Erfindung Polypeptid-Teilsequenzen, sogenannte ORF (open-reading-frame)-Peptide,  
gemäß den Sequenzprotokollen ORF ID Nos. 51-106.

Die Erfindung betrifft ferner die Polypeptid-Sequenzen, die mindestens eine 80%ige Homologie, insbesondere eine  
90%ige Homologie zu den erfindungsgemäßen Polypeptid-Teilsequenzen der ORF ID Nos. 51-106 aufweisen.

Die Erfindung betrifft auch Antikörper, die gegen ein Polypeptid oder Fragment davon gerichtet sind, welche von den  
erfindungsgemäßen Nukleinsäuren der Sequenzen Seq. ID No. 1 bis Seq. ID 50 kodiert werden.

Unter Antikörper sind insbesondere monoklonale und Phage-Display-Antikörper zu verstehen.

Die erfindungsgemäßen Polypeptide der Sequenzen ORF ID Nos. 51-106 können auch als Tool zum Auffinden von  
Wirkstoffen gegen den Blasen-tumor verwendet werden, was ebenfalls Gegenstand der vorliegenden Erfindung ist.

Ebenfalls Gegenstand der vorliegenden Erfindung ist die Verwendung der Nukleinsäure-Sequenzen gemäß den Se-  
quenzen Seq. ID No. 1-50 zur Expression von Polypeptiden, die als Tools zum Auffinden von Wirkstoffen gegen den  
Blasen-tumor verwendet werden können.

Die Erfindung betrifft auch die Verwendung der gefundenen Polypeptid-Teilsequenzen ORF ID No. 51-106 als Arz-  
neimittel in der Gentherapie zur Behandlung gegen den Blasen-tumor, bzw. zur Herstellung eines Arzneimittels zur Be-  
handlung gegen den Blasen-tumor.

Die Erfindung betrifft auch Arzneimittel, die mindestens eine Polypeptid-Teilsequenz ORF ID No. 51-106 enthalten.

Die gefundenen erfindungsgemäßen Nukleinsäure-Sequenzen können auch genomische oder mRNA-Sequenzen sein.

Die Erfindung betrifft auch genomische Gene, ihre Exon- und Intronstruktur und deren Spleißvarianten, erhältlich aus  
den cDNAs der Sequenzen Seq. ID No. 1-50, sowie deren Verwendung zusammen mit geeigneten regulativen Elemen-  
ten, wie geeigneten Promotoren und/oder Enhancern.

Mit den erfindungsgemäßen Nukleinsäuren (cDNA-Sequenzen) Seq. ID No. 1-50 werden genomische BAC-, PAC-  
und Cosmid-Bibliotheken gescreent und über komplementäre Basenpaarung (Hybridisierung) spezifisch humane Klone  
isoliert. Die so isolierten BAC-, PAC- und Cosmid-Klone werden mit Hilfe der Fluoreszenz-in-situ-Hybridisation auf  
Metaphasenchromosomen hybridisiert und entsprechende Chromosomenabschnitte identifiziert, auf denen die entspre-  
chenden genomischen Gene liegen. BAC-, PAC- und Cosmid-Klone werden sequenziert, um die entsprechenden genom-  
ischen Gene in ihrer vollständigen Struktur (Promotoren, Enhancer, Silencer, Exons und Introns) aufzuklären. BAC-,  
PAC- und Cosmid-Klone können als eigenständige Moleküle für den Gentransfer eingesetzt werden (s. Fig. 5).

Die Erfindung betrifft auch BAC-, PAC- und Cosmid-Klone, enthaltend funktionelle Gene und ihre chromosomale  
Lokalisation, entsprechend den Sequenzen Seq. ID. No. 1 bis Seq. ID No. 50, zur Verwendung als Vehikel zum Gentrans-  
fer.

#### Bedeutungen von Fachbegriffen und Abkürzungen

Nukleinsäuren = Unter Nukleinsäuren sind in der vorliegenden Erfindung zu verstehen: mRNA, partielle cDNA, vollän-  
gen cDNA und genomische Gene (Chromosomen).

ORF = Open Reading Frame, eine definierte Abfolge von Aminosäuren, die von der cDNA-Sequenz abgeleitet werden  
kann.

Contig = eine Menge von DNA-Sequenzen, die aufgrund sehr großer Ähnlichkeiten zu einer Sequenz zusammengefaßt  
werden können (Consensus)

Singleton = ein Contig, der nur eine Sequenz enthält.

#### Erklärung zu den Alignmentparametern

minimal initial match = minimaler anfänglicher Identitätsbereich

maximum pads per read = maximale Anzahl von Insertionen

maximum percent mismatch = maximale Abweichung in %.

## Erklärung der Abbildungen

- Fig. 1 zeigt die systematische Gen-Suche in der Incyte LifeSeq Datenbank.  
 Fig. 2a zeigt das Prinzip der EST-Assemblierung  
 Fig. 2b1-2b4 zeigt das gesamte Prinzip der EST-Assemblierung  
 Fig. 3 zeigt die in silico Subtraktion der Genexpression in verschiedenen Geweben  
 Fig. 4a zeigt die Bestimmung der gewebsspezifischen Expression über elektronischen Northern.  
 Fig. 4b zeigt den elektronischen Northern  
 Fig. 5 zeigt die Isolierung von genomischen BAC- und PAC-Klonen.  
 Die nachfolgenden Beispiele erläutern die Herstellung der erfindungsgemäßen Nukleinsäure-Sequenzen, ohne die Erfindung auf diese Beispiele und Nukleinsäure-Sequenzen zu beschränken.

## Beispiel 1

## Suche nach Tumor-bezogenen Kandidatengen

Zuerst wurden sämtliche ESTs des entsprechenden Gewebes aus der LifeSeq-Datenbank (vom Oktober 1997) extrahiert. Diese wurden dann mittels des Programms GAP4 des Staden-Pakets mit den Parametern 0% mismatch, 8 pads per read und einem minimalen match von 20 assembliert. Die nicht in die GAP4-Datenbank aufgenommenen Sequenzen (Fails) wurden erst bei 1% mismatch und dann nochmals bei 2% mismatch mit der Datenbank assembliert. Aus den Contigs der Datenbank, die aus mehr als einer Sequenz bestanden, wurden Consensussequenzen errechnet. Die Singletons der Datenbank, die nur aus einer Sequenz bestanden, wurden mit den nicht in die GAP4-Datenbank aufgenommenen Sequenzen bei 2% mismatch erneut assembliert. Wiederum wurden für die Contigs die Consensussequenzen ermittelt. Alle übrigen ESTs wurden bei 4% mismatch erneut assembliert. Die Consensussequenzen wurden abermals extrahiert und mit den vorherigen Consensussequenzen sowie den Singletons und den nicht in die Datenbank aufgenommenen Sequenzen abschließend bei 4% mismatch assembliert. Die Consensussequenzen wurden gebildet und mit den Singletons und Fails als Ausgangsbasis für die Gewebsvergleiche verwendet. Durch diese Prozedur konnte sichergestellt werden, daß unter den verwendeten Parametern sämtliche Sequenzen von einander unabhängige Genbereiche darstellten.

Fig. 2b1-2b4 veranschaulicht die Verlängerung der Blasengewebe ESTs.  
 Die so assemblierten Sequenzen der jeweiligen Gewebe wurden anschließend mittels des gleichen Programms miteinander verglichen (Fig. 3). Hierzu wurden erst alle Sequenzen des ersten Gewebes in die Datenbank eingegeben. (Daher war es wichtig, daß diese voneinander unabhängig waren.)

Dann wurden alle Sequenzen des zweiten Gewebes mit allen des ersten verglichen. Das Ergebnis waren Sequenzen, die für das erste bzw. das zweite Gewebe spezifisch waren, sowie welche, die in beiden vorkamen. Bei Letzteren wurde das Verhältnis der Häufigkeit des Vorkommens in den jeweiligen Geweben ausgewertet. Sämtliche, die Auswertung der assemblierten Sequenzen betreffenden Programme, wurden selbst entwickelt.

Alle Sequenzen, die mehr als viermal in jeweils einem der verglichenen Gewebe vorkamen, sowie alle, die mindestens fünfmal so häufig in einem der beiden Gewebe vorkamen wurden weiter untersucht. Diese Sequenzen wurden einem elektronischen Northern (s. Beispiel 2.1) unterzogen, wodurch die Verteilung in sämtlichen Tumor- und Normal-Geweben untersucht wurde (s. Fig. 4a und Fig. 4b). Die relevanten Kandidaten wurden dann mit Hilfe sämtlicher Incyte ESTs und allen ESTs öffentlicher Datenbanken verlängert (s. Beispiel 3). Anschließend wurden die Sequenzen und ihre Übersetzung in mögliche Proteine mit allen Nukleotid- und Proteindatenbanken verglichen, sowie auf mögliche, für Proteine kodierende Regionen untersucht.

## Beispiel 2

## Algorithmus zur Identifikation und Verlängerung von partiellen cDNA-Sequenzen mit verändertem Expressionsmuster

Im folgenden soll ein Algorithmus zur Auffindung über- oder unterexprimierter Gene erläutert werden. Die einzelnen Schritte sind der besseren Übersicht halber auch in einem Flußdiagramm zusammengefaßt (s. Fig. 4b).

## 2.1 Elektronischer Northern-Blot

Zu einer partiellen DNA-Sequenz S, z. B. einem einzelnen EST oder einem Contig von ESTs, werden mittels eines Standardprogramms zur Homologiesuche, z. B. BLAST (Altschul, S. F., Gish W., Miller, W., Myers, E. W. und Lipman, D. J. (1990) J. Mol. Biol., 215, 403-410), BLAST2 (Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. und Lipman, D. J. (1997) Nucleic Acids Research 25 3389-3402) oder FASTA (Pearson, W. R. und Lipman, D. J. (1988) Proc. Natl. Acad. Sci. USA 85 2444-2448), die homologen Sequenzen in verschiedenen nach Geweben geordneten (privaten oder öffentlichen) EST-Bibliotheken bestimmt. Die dadurch ermittelten (relativen oder absoluten) Gewebe-spezifischen Vorkommenshäufigkeiten dieser Partial-Sequenz S werden als elektronischer Northern-Blot bezeichnet.

## 2.1.1

Analog der unter 2.1 beschriebenen Verfahrensweise wurde die Sequenz Seq. ID No. 16 gefunden, die 17,7.x stärker im normalen Blasentumorgewebe als im normalem Blasengewebe vorkommt.  
 Das Ergebnis ist wie folgt:

# DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 16

	NORMAL	TUMOR	Verhaeltnisse		
	%Haeufigkeit	%Haeufigkeit	N/T	T/N	
Blase	0.0039	0.0690	0.0565	17.6998	
Brust	0.0000	0.0000	undef	undef	5
Duenndarm	0.0031	0.0000	undef	0.0000	
Eierstock	0.0180	0.0078	2.3025	0.4343	
Endokrines_Gewebe	0.0000	0.0000	undef	undef	
Gastrointestinal	0.0000	0.0000	undef	undef	
Gehirn	0.0000	0.0000	undef	undef	10
Haematopoetisch	0.0000	0.0000	undef	undef	
Haut	0.0000	0.0000	undef	undef	
Hepatisch	0.0000	0.0000	undef	undef	
Herz	0.0000	0.0000	undef	undef	
Hoden	0.0000	0.0000	undef	undef	
Lunge	0.0000	0.0000	undef	undef	15
Magen-Speiserohre	0.0000	0.0000	undef	undef	
Muskel-Skelett	0.0000	0.0000	undef	undef	
Niere	0.0000	0.0000	undef	undef	
Pankreas	0.0000	0.0000	undef	undef	
Penis	0.0000	0.0000	undef	undef	
Prostata	0.0065	0.0064	1.0236	0.9769	20
Uterus_Endometrium	0.0000	0.0000	undef	undef	
Uterus_Myometrium	0.0000	0.0000	undef	undef	
Uterus_allgemein	0.0000	0.0000	undef	undef	
Brust-Hyperplasie	0.0000				
Prostata-Hyperplasie	0.0178				25
Samenblase	0.0000				
Sinnesorgane	0.0000				
Weisse_Blutkoerperchen	0.0000				
Zervix	0.0000				30
FOETUS					
	%Haeufigkeit				
Entwicklung	0.0000				
Gastrointestinal	0.0028				
Gehirn	0.0000				35
Haematopoetisch	0.0000				
Haut	0.0000				
Hepatisch	0.0000				
Herz-Blutgefuesse	0.0000				
Lunge	0.0000				
Nebenniere	0.0000				40
Niere	0.0000				
Placenta	0.0000				
Prostata	0.0000				
Sinnesorgane	0.0000				45
NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN					
	%Haeufigkeit				
Brust	0.0000				
Eierstock_n	0.0000				
Eierstock_t	0.0000				
Endokrines_Gewebe	0.0000				50
Foetal	0.0035				
Gastrointestinal	0.0000				
Haematopoetisch	0.0000				
Haut-Muskel	0.0000				
Hoden	0.0000				55
Lunge	0.0000				
Nerven	0.0000				
Prostata	0.0068				
Sinnesorgane	0.0000				
Uterus_n	0.0000				60

In analoger Verfahrensweise wurden auch folgende Northern gefunden:

65



## DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 8

	NORMAL	TUMOR	Verhaeltnisse	
	%Haeufigkeit	%Haeufigkeit	N/T	T/N
5	Blase 0.0000	0.0281	0.0000	undef
	Brust 0.0038	0.0056	0.6805	1.4694
	Duennndarm 0.0399	0.0000	undef	0.0000
	Eierstock 0.0000	0.0078	0.0000	undef
	Endokrines_Gewebe 0.0000	0.0000	undef	undef
	Gastrointestinal 0.0479	0.0000	undef	0.0000
10	Gehirn 0.0000	0.0010	0.0000	undef
	Haematopoetisch 0.0227	0.0000	undef	0.0000
	Haut 0.0037	0.0000	undef	0.0000
	Hepatisch 0.0000	0.0323	0.0000	undef
	Herz 0.0000	0.0000	undef	undef
	Hoden 0.0000	0.0585	0.0000	undef
15	Lunge 0.0145	0.0123	1.1854	0.8436
	Magen-Speiserohre 0.0000	0.0000	undef	undef
	Muskel-Skelett 0.0120	0.0120	0.9994	1.0006
	Niere 0.0000	0.0274	0.0000	undef
	Pankreas 0.0066	0.0110	0.5983	1.6714
20	Penis 0.0000	0.0000	undef	undef
	Prostata 0.0022	0.0021	1.0236	0.9769
	Uterus_Endometrium 0.0000	0.0000	undef	undef
	Uterus_Myometrium 0.0000	0.0000	undef	undef
	Uterus_allgemein 0.0000	0.0000	undef	undef
	Brust-Hyperplasie 0.0128			
25	Prostata-Hyperplasie 0.0030			
	Samenblase 0.0000			
	Sinnesorgane 0.0000			
	Weisse_Blutkoerperchen 0.0009			
	Zervix 0.0000			

30	FOETUS
	%Haeufigkeit
	Entwicklung 0.0000
	Gastrointestinal 0.0000
35	Gehirn 0.0000
	Haematopoetisch 0.0039
	Haut 0.0000
	Hepatisch 0.0000
	Herz-Blutgefuesse 0.0000
	Lunge 0.0000
40	Nebenniere 0.0000
	Niere 0.0000
	Placenta 0.0000
	Prostata 0.0000
	Sinnesorgane 0.0000

45	NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN
	%Haeufigkeit
	Brust 0.0000
	Eierstock_n 0.0000
50	Eierstock_t 0.0000
	Endokrines_Gewebe 0.0000
	Foetal 0.0000
	Gastrointestinal 0.0000
	Haematopoetisch 0.0000
	Haut-Muskel 0.0000
55	Hoden 0.0000
	Lunge 0.0000
	Nerven 0.0000
	Prostata 0.0000
	Sinnesorgane 0.0000
60	Uterus_n 0.0000

65

# DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 2

	NORMAL	TUMOR	Verhaeltnisse		
	%Haeufigkeit	%Haeufigkeit	N/T	T/N	
Blase	0.0000	0.0307	0.0000	undef	5
Brust	0.0307	0.0376	0.8166	1.2245	
Duennndarm	0.0337	0.0165	2.0391	0.4904	
Eierstock	0.0120	0.0364	0.3289	3.0402	
Endokrines_Gewebe	0.0255	0.0075	3.3962	0.2944	
Gastrointestinal	0.0153	0.0185	0.8283	1.2072	10
Gehirn	0.0185	0.0216	0.8571	1.1667	
Haematopoetisch	0.0201	0.0379	0.5293	1.8892	
Haut	0.0844	0.0000	undef	0.0000	
Hepatisch	0.0238	0.0065	3.6765	0.2720	
Herz	0.0148	0.0000	undef	0.0000	
Hoden	0.0575	0.0351	1.6399	0.6098	15
Lunge	0.0145	0.0082	1.7781	0.5624	
Magen-Speiserohre	0.0387	0.0077	5.0421	0.1983	
Muskel-Skelett	0.0308	0.0300	1.0280	0.9728	
Niere	0.0217	0.0000	undef	0.0000	
Pankreas	0.0099	0.0110	0.8974	1.1143	20
Penis	0.0240	0.0000	undef	0.0000	
Prostata	0.0262	0.0213	1.2284	0.8141	
Uterus_Endometrium	0.0135	0.0000	undef	0.0000	
Uterus_Myometrium	0.0152	0.0408	0.3741	2.6732	
Uterus_allgemein	0.0204	0.0954	0.2135	4.6839	25
Brust-Hyperplasie	0.0512				
Prostata-Hyperplasie	0.0268				
Samenblase	0.0089				
Sinnesorgane	0.0235				
Weisse_Blutkoerperchen	0.0286				30
Zervix	0.0106				

	FOETUS	
	%Haeufigkeit	
Entwicklung	0.0278	35
Gastrointestinal	0.0305	
Gehirn	0.0063	
Haematopoetisch	0.0157	
Haut	0.0000	40
Hepatisch	0.0000	
Herz-Blutgefuesse	0.0213	
Lunge	0.0289	
Nebenniere	0.0000	
Niere	0.0185	45
Placenta	0.0121	
Prostata	0.0000	
Sinnesorgane	0.0126	

	NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN	
	%Haeufigkeit	
Brust	0.0204	
Eierstock_n	0.0000	
Eierstock_t	0.0051	55
Endokrines_Gewebe	0.0000	
Foetal	0.0122	
Gastrointestinal	0.0488	
Haematopoetisch	0.0000	
Haut-Muskel	0.0000	60
Hoden	0.0463	
Lunge	0.0164	
Nerven	0.0100	
Prostata	0.0137	
Sinnesorgane	0.0000	65
Uterus_n	0.0125	

## DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 3

		NORMAL	TUMOR	Verhaeltnisse	
		%Haefufigkeit	%Haefufigkeit	N/T	T/N
5	Blase	0.0000	0.0256	0.0000	undef
	Brust	0.0000	0.0000	undef	undef
	Duennndarm	0.0000	0.0000	undef	undef
	Eierstock	0.0000	0.0000	undef	undef
	Endokrines_Gewebe	0.0000	0.0000	undef	undef
10	Gastrointestinal	0.0000	0.0000	undef	undef
	Gehirn	0.0000	0.0000	undef	undef
	Haematopoetisch	0.0000	0.0000	undef	undef
	Haut	0.0000	0.0000	undef	undef
	Hepatisch	0.0000	0.0000	undef	undef
15	Herz	0.0000	0.0000	undef	undef
	Hoden	0.0000	0.0000	undef	undef
	Lunge	0.0000	0.0000	undef	undef
	Magen-Speiserohre	0.0000	0.0000	undef	undef
	Muskel-Skelett	0.0000	0.0000	undef	undef
20	Niere	0.0000	0.0000	undef	undef
	Pankreas	0.0000	0.0000	undef	undef
	Penis	0.0030	0.0000	undef	0.0000
	Prostata	0.0000	0.0000	undef	undef
	Uterus_Endometrium	0.0000	0.0000	undef	undef
	Uterus_Myometrium	0.0000	0.0000	undef	undef
25	Uterus_allgemein	0.0000	0.0000	undef	undef
	Brust-Hyperplasie	0.0000			
	Prostata-Hyperplasie	0.0000			
	Samenblase	0.0000			
	Sinnesorgane	0.0000			
30	Weisse_Blutkoerperchen	0.0000			
	Zervix	0.0000			

		FOETUS
		%Haefufigkeit
35	Entwicklung	0.0000
	Gastrointestinal	0.0000
	Gehirn	0.0000
	Haematopoetisch	0.0000
40	Haut	0.0000
	Hepatisch	0.0000
	Herz-Blutgefuesse	0.0000
	Lunge	0.0000
	Nebenniere	0.0000
45	Niere	0.0000
	Placenta	0.0000
	Prostata	0.0000
	Sinnesorgane	0.0000

		NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN
		%Haefufigkeit
50	Brust	0.0000
	Eierstock_n	0.0000
	Eierstock_t	0.0000
55	Endokrines_Gewebe	0.0000
	Foetal	0.0000
	Gastrointestinal	0.0000
	Haematopoetisch	0.0000
60	Haut-Muskel	0.0000
	Hoden	0.0000
	Lunge	0.0000
	Nerven	0.0000
	Prostata	0.0000
65	Sinnesorgane	0.0000
	Uterus_n	0.0000

## DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 4

	NORMAL %Haeufigkeit	TUMOR %Haeufigkeit	Verhaeltnisse		
			N/T	T/N	
Blase	0.0000	0.0230	0.0000	undef	5
Brust	0.0090	0.0094	0.9527	1.0496	
Duenndarm	0.0123	0.0165	0.7415	1.3487	
Eierstock	0.0150	0.0104	1.4391	0.6949	
Endokrines_Gewebe	0.0068	0.0125	0.5434	1.8403	10
Gastrointestinal	0.0038	0.0046	0.8283	1.2072	
Gehirn	0.0044	0.0123	0.3600	2.7779	
Haematopoetisch	0.0187	0.0000	undef	0.0000	
Haut	0.0000	0.0000	undef	undef	15
Hepatisch	0.0000	0.0388	0.0000	undef	
Herz	0.0138	0.0000	undef	0.0000	
Hoden	0.0173	0.0000	undef	0.0000	
Lunge	0.0093	0.0143	0.6532	1.5310	20
Magen-Speiserohre	0.0097	0.0000	undef	0.0000	
Muskel-Skelett	0.0017	0.0180	0.0952	10.5060	
Niere	0.0136	0.0137	0.9913	1.0088	
Pankreas	0.0149	0.0055	2.6923	0.3714	25
Penis	0.0150	0.0000	undef	0.0000	
Prostata	0.0065	0.0064	1.0236	0.9769	
Uterus_Endometrium	0.0000	0.0000	undef	undef	
Uterus_Myometrium	0.0000	0.0136	0.0000	undef	30
Uterus_allgemein	0.0000	0.0000	undef	undef	
Brust-Hyperplasie	0.0032				
Prostata-Hyperplasie	0.0178				
Samenblase	0.0089				35
Sinnesorgane	0.0000				
Weisse_Blutkoerperchen	0.0035				
Zervix	0.0000				

FOETUS				
	%Haeufigkeit			
Entwicklung	0.0000			35
Gastrointestinal	0.0167			
Gehirn	0.0000			
Haematopoetisch	0.0354			
Haut	0.0000			40
Hepatisch	0.0000			
Herz-Blutgefuesse	0.0000			
Lunge	0.0072			
Nebenniere	0.0000			45
Niere	0.0062			
Placenta	0.0000			
Prostata	0.0000			
Sinnesorgane	0.0000			50

NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN				
	%Haeufigkeit			
Brust	0.0272			55
Eierstock_n	0.0000			
Eierstock_t	0.0051			
Endokrines_Gewebe	0.0000			
Foetal	0.0151			60
Gastrointestinal	0.0000			
Haematopoetisch	0.0114			
Haut-Muskel	0.0259			
Hoden	0.0077			65
Lunge	0.0000			
Nerven	0.0060			
Prostata	0.0068			
Sinnesorgane	0.0155			65
Uterus_n	0.0250			

## DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 5

		NORMAL	TUMOR	Verhaeltnisse	
		%Haeufigkeit	%Haeufigkeit	N/T	T/N
5	Blase	0.0000	0.0204	0.0000	undef
	Brust	0.0026	0.0056	0.4537	2.2042
	Duennndarm	0.0031	0.0000	undef	0.0000
	Eierstock	0.0060	0.0156	0.3838	2.6058
	Endokrines_Gewebe	0.0051	0.0000	undef	0.0000
10	Gastrointestinal	0.0134	0.0093	1.4496	0.6898
	Gehirn	0.0052	0.0051	1.0079	0.9921
	Haematopoetisch	0.0000	0.0000	undef	undef
	Haut	0.0000	0.0000	undef	undef
	Hepatisch	0.0000	0.0000	undef	undef
15	Herz	0.0032	0.0000	undef	0.0000
	Hoden	0.0000	0.0000	undef	undef
	Lunge	0.0031	0.0020	1.5241	0.6561
	Magen-Speiserohre	0.0000	0.0000	undef	undef
	Muskel-Skelett	0.0034	0.0060	0.5711	1.7510
20	Niere	0.0190	0.0068	2.7756	0.3603
	Pankreas	0.0017	0.0000	undef	0.0000
	Penis	0.0000	0.0000	undef	undef
	Prostata	0.0131	0.0043	3.0709	0.3256
	Uterus_Endometrium	0.0068	0.0000	undef	0.0000
25	Uterus_Myometrium	0.0000	0.0000	undef	undef
	Uterus_allgemein	0.0000	0.0000	undef	undef
	Brust-Hyperplasie	0.0000			
	Prostata-Hyperplasie	0.0059			
	Samenblase	0.0000			
	Sinnesorgane	0.0000			
30	Weisse_Blutkoerperchen	0.0000			
	Zervix	0.0000			

		FOETUS
		%Haeufigkeit
35	Entwicklung	0.0000
	Gastrointestinal	0.0083
	Gehirn	0.0063
	Haematopoetisch	0.0079
40	Haut	0.0000
	Hepatisch	0.0000
	Herz-Blutgefuesse	0.0000
	Lunge	0.0000
	Nebenniere	0.0000
45	Niere	0.0000
	Placenta	0.0000
	Prostata	0.0000
	Sinnesorgane	0.0000

		NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN
		%Haeufigkeit
50	Brust	0.0000
	Eierstock_n	0.0000
55	Eierstock_t	0.0051
	Endokrines_Gewebe	0.0000
	Foetal	0.0000
	Gastrointestinal	0.0000
	Haematopoetisch	0.0000
60	Haut-Muskel	0.0000
	Hoden	0.0000
	Lunge	0.0082
	Nerven	0.0050
	Prostata	0.0137
65	Sinnesorgane	0.0000
	Uterus_n	0.0000

# DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 6

	NORMAL	TUMOR	Verhaeltnisse		
	%Haeufigkeit	%Haeufigkeit	N/T	T/N	
Blase	0.0000	0.0204	0.0000	undef	5
Brust	0.0038	0.0075	0.5104	1.9593	
Duennndarm	0.0031	0.0000	undef	0.0000	
Eierstock	0.0060	0.0052	1.1513	0.8686	
Endokrines_Gewebe	0.0017	0.0000	undef	0.0000	
Gastrointestinal	0.0038	0.0000	undef	0.0000	10
Gehirn	0.0022	0.0021	1.0799	0.9260	
Haematopoetisch	0.0053	0.0000	undef	0.0000	
Haut	0.0000	0.0000	undef	undef	
Hepatisch	0.0048	0.0000	undef	0.0000	
Herz	0.0042	0.0137	0.3084	3.2426	15
Hoden	0.0000	0.0000	undef	undef	
Lunge	0.0052	0.0143	0.3629	2.7557	
Magen-Speiserohre	0.0000	0.0000	undef	undef	
Muskel-Skelett	0.0017	0.0060	0.2856	3.5020	
Niere	0.0000	0.0000	undef	undef	20
Pankreas	0.0017	0.0110	0.1496	6.6857	
Penis	0.0000	0.0000	undef	undef	
Prostata	0.0000	0.0021	0.0000	undef	
Uterus_Endometrium	0.0135	0.0000	undef	0.0000	
Uterus_Myometrium	0.0229	0.0068	3.3668	0.2970	25
Uterus_allgemein	0.0204	0.0000	undef	0.0000	
Brust-Hyperplasie	0.0000				
Prostata-Hyperplasie	0.0000				
Samenblase	0.0000				
Sinnesorgane	0.0353				30
Weisse_Blutkoerperchen	0.0009				
Zervix	0.0000				

FOETUS		
	%Haeufigkeit	
Entwicklung	0.0000	35
Gastrointensteinal	0.0000	
Gehirn	0.0000	
Haematopoetisch	0.0118	40
Haut	0.0000	
Hepatisch	0.0000	
Herz-Blutgefaesse	0.0000	
Lunge	0.0036	
Nebenniere	0.0000	45
Niere	0.0124	
Placenta	0.0242	
Prostata	0.0000	
Sinnesorgane	0.0000	

NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN			50
%Haeufigkeit			
Brust	0.0000		
Eierstock_n	0.0000		
Eierstock_t	0.0000		
Endokrines_Gewebe	0.0245		55
Foetal	0.0151		
Gastrointestinal	0.0000		
Haematopoetisch	0.0000		
Haut-Muskel	0.0259		
Hoden	0.0000		60
Lunge	0.0164		
Nerven	0.0080		
Prostata	0.0000		
Sinnesorgane	0.0000		
Uterus_n	0.0250		65

## DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 7

	NORMAL	TUMOR	Verhaeltnisse	
	%Haeufigkeit	%Haeufigkeit	N/T	T/N
5	Blase 0.0000	0.0179	0.0000	undef
	Brust 0.0038	0.0132	0.2917	3.4287
	Duennndarm 0.0061	0.0000	undef	0.0000
	Eierstock 0.0060	0.0052	1.1513	0.8686
	Endokrines_Gewebe 0.0017	0.0100	0.1698	5.8889
10	Gastrointestinal 0.0019	0.0370	0.0518	19.3158
	Gehirn 0.0096	0.0051	1.8719	0.5342
	Haematopoetisch 0.0067	0.0000	undef	0.0000
	Haut 0.0000	0.0000	undef	undef
	Hepatisch 0.0048	0.0065	0.7353	1.3600
15	Herz 0.0138	0.0137	1.0023	0.9977
	Hoden 0.0288	0.0468	0.6150	1.6261
	Lunge 0.0031	0.0143	0.2177	4.5929
	Magen-Speiserohre 0.0387	0.0153	2.5211	0.3967
	Muskel-Skelett 0.0017	0.0000	undef	0.0000
20	Niere 0.0054	0.0000	undef	0.0000
	Pankreas 0.0050	0.0055	0.8974	1.1143
	Penis 0.0000	0.0000	undef	undef
	Prostata 0.0000	0.0021	0.0000	undef
	Uterus_Endometrium 0.0000	0.0000	undef	undef
	Uterus_Myometrium 0.0000	0.0000	undef	undef
25	Uterus_allgemein 0.0102	0.0000	undef	0.0000
	Brust-Hyperplasie 0.0000			
	Prostata-Hyperplasie 0.0000			
	Samenblase 0.0000			
	Sinnesorgane 0.0000			
30	Weisse_Blutkoerperchen 0.0026			
	Zervix 0.0000			

	FOETUS
	%Haeufigkeit
35	Entwicklung 0.0000
	Gastrointestinal 0.0028
	Gehirn 0.0751
	Haematopoetisch 0.0079
40	Haut 0.0000
	Hepatisch 0.0000
	Herz-Blutgefuesse 0.0036
	Lunge 0.0036
	Nebenniere 0.0000
45	Niere 0.0000
	Placenta 0.0061
	Prostata 0.0499
	Sinnesorgane 0.0000

	NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN
	%Haeufigkeit
50	Brust 0.0068
	Eierstock_n 0.0000
55	Eierstock_t 0.0000
	Endokrines_Gewebe 0.0000
	Foetal 0.0087
	Gastrointestinal 0.0122
	Haematopoetisch 0.0114
60	Haut-Muskel 0.0097
	Hoden 0.0540
	Lunge 0.0082
	Nerven 0.0201
	Prostata 0.0205
65	Sinnesorgane 0.0000
	Uterus_n 0.0375

# DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 8

	NORMAL	TUMOR	Verhaeltnisse		
	%Haeufigkeit	%Haeufigkeit	N/T	T/N	
Blase	0.0000	0.0179	0.0000	undef	5
Brust	0.0013	0.0038	0.3403	2.9389	
Duennndarm	0.0123	0.0000	undef	0.0000	
Eierstock	0.0060	0.0078	0.7675	1.3029	
Endokrines Gewebe	0.0068	0.0100	0.6792	1.4722	
Gastrointestinal	0.0038	0.0093	0.4142	2.4145	10
Gehirn	0.0044	0.0144	0.3086	3.2409	
Haematopoetisch	0.0080	0.0000	undef	0.0000	
Haut	0.0037	0.0000	undef	0.0000	
Hepatisch	0.0000	0.0000	undef	undef	
Herz	0.0021	0.0000	undef	0.0000	15
Hoden	0.0058	0.0000	undef	0.0000	
Lunge	0.0031	0.0082	0.3810	2.6245	
Magen-Speiserohre	0.0000	0.0307	0.0000	undef	
Muskel-Skelett	0.0051	0.0060	0.8567	1.1673	
Niere	0.0000	0.0000	undef	undef	20
Pankreas	0.0066	0.0000	undef	0.0000	
Penis	0.0060	0.0000	undef	0.0000	
Prostata	0.0087	0.0064	1.3648	0.7327	
Uterus_Endometrium	0.0000	0.0000	undef	undef	
Uterus_Myometrium	0.0000	0.0068	0.0000	undef	25
Uterus_allgemein	0.0051	0.0000	undef	0.0000	
Brust-Hyperplasie	0.0000				
Prostata-Hyperplasie	0.0059				
Samenblase	0.0000				
Sinnesorgane	0.0000				30
Weisse_Blutkoerperchen	0.0035				
Zervix	0.0000				
					35
					40
					45
					50
					55
					60
					65



## DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 9

		NORMAL	TUMOR	Verhaeltnisse	
		%Haeufigkeit	%Haeufigkeit	N/T	T/N
5	Blase	0.0000	0.0179	0.0000	undef
	Brust	0.0064	0.0056	1.1342	0.8817
	Duenn darm	0.0000	0.0165	0.0000	undef
	Eierstock	0.0030	0.0026	1.1513	0.8686
	Endokrines Gewebe	0.0153	0.0025	6.1132	0.1636
10	Gastrointestinal	0.0038	0.0046	0.8283	1.2072
	Gehirn	0.0015	0.0051	0.2880	3.4724
	Haematopoetisch	0.0067	0.0000	undef	0.0000
	Haut	0.0037	0.0847	0.0433	23.0839
	Hepatisch	0.0000	0.0065	0.0000	undef
15	Herz	0.0053	0.0000	undef	0.0000
	Hoden	0.0230	0.0234	0.9839	1.0163
	Lunge	0.0042	0.0041	1.0161	0.9842
	Magen-Speiserohre	0.0000	0.0077	0.0000	undef
	Muskel-Skelett	0.0034	0.0000	undef	0.0000
20	Niere	0.0054	0.0068	0.7930	1.2610
	Pankreas	0.0017	0.0000	undef	0.0000
	Penis	0.0030	0.0267	0.1123	8.9035
	Prostata	0.0065	0.0064	1.0236	0.9769
	Uterus_Endometrium	0.0135	0.0000	undef	0.0000
25	Uterus_Myometrium	0.0152	0.0000	undef	0.0000
	Uterus_allgemein	0.0000	0.0000	undef	undef
	Brust-Hyperplasie	0.0032			
	Prostata-Hyperplasie	0.0000			
	Samenblase	0.0000			
30	Sinnesorgane	0.0000			
	Weisse_Blutkoerperchen	0.0043			
	Zervix	0.0000			

		FOETUS
		%Haeufigkeit
35	Entwicklung	0.0000
	Gastrointestinal	0.0000
	Gehirn	0.0000
	Haematopoetisch	0.0039
40	Haut	0.0000
	Hepatisch	0.0000
	Herz-Blutgefasse	0.0036
	Lunge	0.0108
	Nebenniere	0.0000
45	Niere	0.0062
	Placenta	0.0000
	Prostata	0.0499
	Sinnesorgane	0.0000

		NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN
		%Haeufigkeit
50	Brust	0.0068
	Eierstock_n	0.0000
55	Eierstock_t	0.0101
	Endokrines_Gewebe	0.0000
	Foetal	0.0047
	Gastrointestinal	0.0000
	Haematopoetisch	0.0114
60	Haut-Muskel	0.0194
	Hoden	0.0309
	Lunge	0.0000
	Nerven	0.0040
	Prostata	0.0000
65	Sinnesorgane	0.0000
	Uterus_n	0.0125

# DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 10

	NORMAL	TUMOR	Verhaeltnisse		
	%Haeufigkeit	%Haeufigkeit	N/T	T/N	
Blase	0.0000	0.0153	0.0000	undef	5
Brust	0.0051	0.0019	2.7221	0.3674	
Duennndarm	0.0000	0.0000	undef	undef	
Eierstock	0.0000	0.0000	undef	undef	
Endokrines Gewebe	0.0000	0.0000	undef	undef	
Gastrointestinal	0.0000	0.0093	0.0000	undef	10
Gehirn	0.0000	0.0113	0.0000	undef	
Haematopoetisch	0.0000	0.0000	undef	undef	
Haut	0.0000	0.0000	undef	undef	
Hepatisch	0.0000	0.0129	0.0000	undef	
Herz	0.0011	0.0000	undef	0.0000	15
Hoden	0.0058	0.0000	undef	0.0000	
Lunge	0.0021	0.0000	undef	0.0000	
Magen-Speiserohre	0.0000	0.0000	undef	undef	
Muskel-Skelett	0.0000	0.0000	undef	undef	
Niere	0.0000	0.0000	undef	undef	20
Pankreas	0.0000	0.0000	undef	undef	
Penis	0.0000	0.0000	undef	undef	
Prostata	0.0000	0.0000	undef	undef	
Uterus_Endometrium	0.0000	0.0000	undef	undef	
Uterus_Myometrium	0.0000	0.0000	undef	undef	25
Uterus_allgemein	0.0000	0.0000	undef	undef	
Brust-Hyperplasie	0.0000				
Prostata-Hyperplasie	0.0000				
Samenblase	0.0000				
Sinnesorgane	0.0000				30
Weisse_Blutkoerperchen	0.0000				
Zervix	0.0000				
	FOETUS				
	%Haeufigkeit				35
Entwicklung	0.0000				
Gastrointestinal	0.0000				
Gehirn	0.0000				
Haematopoetisch	0.0000				
Haut	0.0000				40
Hepatisch	0.0000				
Herz-Blutgefuesse	0.0036				
Lunge	0.0000				
Nebenniere	0.0000				
Niere	0.0000				45
Placenta	0.0000				
Prostata	0.0000				
Sinnesorgane	0.0000				
	NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN				50
	%Haeufigkeit				
Brust	0.0000				
Eierstock_n	0.0000				
Eierstock_t	0.0000				55
Endokrines_Gewebe	0.0000				
Foetal	0.0006				
Gastrointestinal	0.0000				
Haematopoetisch	0.0000				
Haut-Muskel	0.0000				60
Hoden	0.0000				
Lunge	0.0000				
Nerven	0.0010				
Prostata	0.0000				
Sinnesorgane	0.0000				65
Uterus_n	0.0000				

## DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 11

	NORMAL	TUMOR	Verhaeltnisse	
	%Haeufigkeit	%Haeufigkeit	N/T	T/N
5	Blase 0.0000	0.0153	0.0000	undef
	Brust 0.0038	0.0038	1.0208	0.9796
	Duenndarm 0.0000	0.0000	undef	undef
	Eierstock 0.0000	0.0104	0.0000	undef
	Endokrines Gewebe 0.0068	0.0025	2.7170	0.3681
10	Gastrointestinal 0.0077	0.0093	0.8283	1.2072
	Gehirn 0.0000	0.0062	0.0000	undef
	Haematopoetisch 0.0107	0.0000	undef	0.0000
	Haut 0.0073	0.0000	undef	0.0000
	Hepatisch 0.0000	0.0194	0.0000	undef
15	Herz 0.0021	0.0000	undef	0.0000
	Hoden 0.0000	0.0000	undef	undef
	Lunge 0.0010	0.0041	0.2540	3.9367
	Magen-Speiserohre 0.0097	0.0077	1.2605	0.7933
	Muskel-Skelett 0.0000	0.0000	undef	undef
20	Niere 0.0054	0.0068	0.7930	1.2610
	Pankreas 0.0050	0.0000	undef	0.0000
	Penis 0.0000	0.0267	0.0000	undef
	Prostata 0.0065	0.0298	0.2193	4.5590
	Uterus_Endometrium 0.0068	0.0528	0.1280	7.8106
25	Uterus_Myometrium 0.0076	0.0068	1.1223	0.8911
	Uterus_allgemein 0.0000	0.0000	undef	undef
	Brust-Hyperplasie 0.0000			
	Prostata-Hyperplasie 0.0238			
	Samenblase 0.0000			
	Sinnesorgane 0.0000			
30	Weisse_Blutkoerperchen 0.0026			
	Zervix 0.0106			

	FOETUS
	%Haeufigkeit
35	Entwicklung 0.0139
	Gastrointestinal 0.0194
	Gehirn 0.0063
	Haematopoetisch 0.0000
40	Haut 0.0000
	Hepatisch 0.0260
	Herz-Blutgefuesse 0.0071
	Lunge 0.0108
	Nebenniere 0.0000
45	Niere 0.0000
	Placenta 0.0121
	Prostata 0.0499
	Sinnesorgane 0.0000

	NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN
	%Haeufigkeit
50	Brust 0.0000
	Eierstock_n 0.0000
55	Eierstock_t 0.0557
	Endokrines_Gewebe 0.0000
	Foetal 0.0076
	Gastrointestinal 0.0000
	Haematopoetisch 0.0000
60	Haut-Muskel 0.0000
	Hoden 0.0000
	Lunge 0.0000
	Nerven 0.0030
	Prostata 0.0137
65	Sinnesorgane 0.0387
	Uterus_n 0.0042

# DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 12

	NORMAL	TUMOR	Verhaeltnisse	
	%Haeufigkeit	%Haeufigkeit	N/T T/N	
Blase	0.0000	0.0153	0.0000 undef	5
Brust	0.0013	0.0000	undef 0.0000	
Duennndarm	0.0031	0.0000	undef 0.0000	
Eierstock	0.0000	0.0026	0.0000 undef	
Endokrines_Gewebe	0.0017	0.0075	0.2264 4.4166	
Gastrointestinal	0.0000	0.0093	0.0000 undef	10
Gehirn	0.0044	0.0000	undef 0.0000	
Haematopoetisch	0.0000	0.0000	undef undef	
Haut	0.0037	0.0000	undef 0.0000	
Hepatisch	0.0000	0.0000	undef undef	
Herz	0.0000	0.0000	undef undef	
Hoden	0.0000	0.0117	0.0000 undef	15
Lunge	0.0000	0.0000	undef undef	
Magen-Speiseroehre	0.0000	0.0000	undef undef	
Muskel-Skelett	0.0000	0.0000	undef undef	
Niere	0.0000	0.0000	undef undef	
Pankreas	0.0017	0.0000	undef 0.0000	20
Penis	0.0000	0.0000	undef undef	
Prostata	0.0000	0.0021	0.0000 undef	
Uterus_Endometrium	0.0000	0.0000	undef undef	
Uterus_Myometrium	0.0000	0.0000	undef undef	
Uterus_allgemein	0.0000	0.0000	undef undef	25
Brust-Hyperplasie	0.0032			
Prostata-Hyperplasie	0.0000			
Samenblase	0.0000			
Sinnesorgane	0.0000			
Weisse_Blutkoerperchen	0.0000			30
Zervix	0.0000			

	FOETUS	
	%Haeufigkeit	
Entwicklung	0.0000	35
Gastrointestinal	0.0000	
Gehirn	0.0125	
Haematopoetisch	0.0000	
Haut	0.0000	40
Hepatisch	0.0000	
Herz-Blutgefuesse	0.0000	
Lunge	0.0036	
Nebenniere	0.0000	
Niere	0.0000	45
Placenta	0.0000	
Prostata	0.0000	
Sinnesorgane	0.0000	

	NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN	
	%Haeufigkeit	
Brust	0.0000	
Eierstock_n	0.0000	
Eierstock_t	0.0000	
Endokrines_Gewebe	0.0245	55
Foetal	0.0023	
Gastrointestinal	0.0000	
Haematopoetisch	0.0000	
Haut-Muskel	0.0000	60
Hoden	0.0000	
Lunge	0.0082	
Nerven	0.0020	
Prostata	0.0000	
Sinnesorgane	0.0000	
Uterus_n	0.0000	65

## DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 13

		NORMAL	TUMOR	Verhaeltnisse	
		%Haeufigkeit	%Haeufigkeit	N/T	T/N
5	Blase	0.0000	0.0153	0.0000	undef
	Brust	0.0090	0.0169	0.5293	1.8893
	Duenn darm	0.0031	0.0000	undef	0.0000
	Eierstock	0.0060	0.0052	1.1513	0.8686
	Endokrines Gewebe	0.0051	0.0226	0.2264	4.4166
10	Gastrointestinal	0.0057	0.0231	0.2485	4.0241
	Gehirn	0.0052	0.0082	0.6300	1.5874
	Haematopoetisch	0.0107	0.0000	undef	0.0000
	Haut	0.0037	0.0000	undef	0.0000
	Hepatisch	0.0000	0.0000	undef	undef
15	Herz	0.0138	0.0000	undef	0.0000
	Hoden	0.0345	0.0117	2.9518	0.3388
	Lunge	0.0021	0.0123	0.1693	5.9051
	Magen-Speiserohre	0.0000	0.0000	undef	undef
	Muskel-Skelett	0.0051	0.0060	0.8567	1.1673
20	Niere	0.0163	0.0068	2.3791	0.4203
	Pankreas	0.0000	0.0055	0.0000	undef
	Penis	0.0030	0.0267	0.1123	8.9035
	Prostata	0.0174	0.0128	1.3648	0.7327
	Uterus_Endometrium	0.0068	0.0000	undef	0.0000
	Uterus_Myometrium	0.0000	0.0000	undef	undef
25	Uterus_allgemein	0.0051	0.0000	undef	0.0000
	Brust-Hyperplasie	0.0000			
	Prostata-Hyperplasie	0.0149			
	Samenblase	0.0267			
	Sinnesorgane	0.0118			
30	Weisse_Blutkoerperchen	0.0087			
	Zervix	0.0000			

		FOETUS
		%Haeufigkeit
35	Entwicklung	0.0139
	Gastrointestinal	0.0139
	Gehirn	0.0063
	Haematopoetisch	0.0079
40	Haut	0.0000
	Hepatisch	0.0000
	Herz-Blutgefuesse	0.0107
	Lunge	0.0072
	Nebenniere	0.0000
45	Niere	0.0000
	Placenta	0.0061
	Prostata	0.0249
	Sinnesorgane	0.0000

		NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN
		%Haeufigkeit
50	Brust	0.0000
	Eierstock_n	0.0000
55	Eierstock_t	0.0101
	Endokrines_Gewebe	0.0000
	Foetal	0.0146
	Gastrointestinal	0.0000
	Haematopoetisch	0.0057
60	Haut-Muskel	0.0000
	Hoden	0.0309
	Lunge	0.0000
	Nerven	0.0010
	Prostata	0.0137
65	Sinnesorgane	0.0000
	Uterus_n	0.0083

## DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 14

	NORMAL	TUMOR	Verhaeltnisse	
	%Haeufigkeit	%Haeufigkeit	N/T T/N	
Blase	0.0000	0.0153	0.0000 undef	5
Brust	0.0051	0.0395	0.1296 7.7146	
Duendarm	0.0000	0.0000	undef undef	
Eierstock	0.0060	0.0234	0.2558 3.9088	
Endokrines_Gewebe	0.0000	0.0000	undef undef	
Gastrointestinal	0.0115	0.0000	undef 0.0000	10
Gehirn	0.0000	0.0041	0.0000 undef	
Haematopoetisch	0.0027	0.0000	undef 0.0000	
Haut	0.0037	0.0000	undef 0.0000	
Hepatisch	0.0381	0.0129	2.9412 0.3400	
Herz	0.0021	0.0000	undef 0.0000	15
Hoden	0.0000	0.0234	0.0000 undef	
Lunge	0.0073	0.0245	0.2964 3.3743	
Magen-Speiserohre	0.0000	0.0537	0.0000 undef	
Muskel-Skelett	0.0137	0.0420	0.3263 3.0643	
Niere	0.0163	0.0000	undef 0.0000	20
Pankreas	0.0017	0.0331	0.0499 20.0570	
Penis	0.0000	0.0267	0.0000 undef	
Prostata	0.0022	0.0128	0.1706 5.8615	
Uterus_Endometrium	0.0068	0.0000	undef 0.0000	
Uterus_Myometrium	0.0305	0.0136	2.2445 0.4455	25
Uterus_allgemein	0.0000	0.0000	undef undef	
Brust-Hyperplasie	0.0000			
Prostata-Hyperplasie	0.0000			
Samenblase	0.0000			
Sinnesorgane	0.0118			
Weisse_Blutkoerperchen	0.0000			30
Zervix	0.0319			

	FOETUS	
	%Haeufigkeit	
Entwicklung	0.0278	35
Gastrointestinal	0.0583	
Gehirn	0.0000	
Haematopoetisch	0.0354	
Haut	0.5025	40
Hepatisch	0.0000	
Herz-Blutgefuesse	0.0071	
Lunge	0.0542	
Nebenniere	0.0761	
Niere	0.1235	45
Placenta	0.0727	
Prostata	0.0249	
Sinnesorgane	0.1255	

	NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN	
	%Haeufigkeit	
Brust	0.0204	
Eierstock_n	0.0000	
Eierstock_t	0.0709	55
Endokrines_Gewebe	0.0245	
Foetal	0.0175	
Gastrointestinal	0.0122	
Haematopoetisch	0.0000	
Haut-Muskel	0.0162	60
Hoden	0.0000	
Lunge	0.0000	
Nerven	0.0000	
Prostata	0.0000	
Sinnesorgane	0.0000	
Uterus_n	0.0000	65

## DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 15

		NORMAL	TUMOR	Verhaeltnisse	
		%Haefufigkeit	%Haefufigkeit	N/T	T/N
5	Blase	0.0039	0.1278	0.0305	32.7774
	Brust	0.0179	0.0414	0.4331	2.3091
	Duenn darm	0.0031	0.1323	0.0232	43.1571
	Eierstock	0.0090	0.0234	0.3838	2.6058
	Endokrines Gewebe	0.0358	0.0301	1.1887	0.8413
10	Gastrointestinal	0.0019	0.0000	undef	0.0000
	Gehirn	0.0007	0.0010	0.7200	1.3890
	Haematopoetisch	0.0000	0.0000	undef	undef
	Haut	0.0037	0.0000	undef	0.0000
	Hepatisch	0.0000	0.0065	0.0000	undef
15	Herz	0.0244	0.0000	undef	0.0000
	Hoden	0.0000	0.0000	undef	undef
	Lunge	0.0374	0.0491	0.7621	1.3122
	Magen-Speiserohre	0.0290	0.0077	3.7816	0.2644
	Muskel-Skelett	0.0000	0.0000	undef	undef
20	Niere	0.0027	0.0068	0.3965	2.5219
	Pankreas	0.0017	0.2209	0.0075	133.7133
	Penis	0.0000	0.0000	undef	undef
	Prostata	0.0065	0.0085	0.7677	1.3026
	Uterus Endometrium	0.0000	0.0000	undef	undef
25	Uterus Myometrium	0.0152	0.0000	undef	0.0000
	Uterus_allgemein	0.0000	0.0000	undef	undef
	Brust-Hyperplasie	0.0192			
	Prostata-Hyperplasie	0.0030			
	Samenblase	0.0267			
30	Sinnesorgane	0.0000			
	Weisse_Blutkoerperchen	0.0000			
	Zervix	0.0106			

		FOETUS
		%Haefufigkeit
35	Entwicklung	0.0000
	Gastrointestinal	0.0000
	Gehirn	0.0000
	Haematopoetisch	0.0000
40	Haut	0.0000
	Hepatisch	0.0000
	Herz-Blutgefuesse	0.0000
	Lunge	0.0000
	Nebenniere	0.0000
45	Niere	0.0000
	Placenta	0.0667
	Prostata	0.0000
	Sinnesorgane	0.0000

		NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN
		%Haefufigkeit
50	Brust	0.1224
	Eierstock_n	0.0000
55	Eierstock_t	0.1013
	Endokrines_Gewebe	0.0000
	Foetal	0.0268
	Gastrointestinal	0.0122
	Haematopoetisch	0.0057
60	Haut-Muskel	0.0065
	Hoden	0.0000
	Lunge	0.0246
	Nerven	0.0000
	Prostata	0.0205
65	Sinnesorgane	0.0000
	Uterus_n	0.0125

# DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 17

	NORMAL %Haeufigkeit	TUMOR %Haeufigkeit	Verhaeltnisse		
			N/T	T/N	
Blase	0.0039	0.0537	0.0726	13.7665	5
Brust	0.0077	0.0207	0.3712	2.6940	
Duenn darm	0.0368	0.0000	undef	0.0000	
Eierstock	0.0150	0.0078	1.9188	0.5212	
Endokrines Gewebe	0.0102	0.0100	1.0189	0.9815	
Gastrointestinal	0.0421	0.0093	4.5559	0.2195	10
Gehirn	0.0118	0.0195	0.6063	1.6494	
Haematopoetisch	0.0174	0.0379	0.4587	2.1798	
Haut	0.0110	0.0000	undef	0.0000	
Hepatisch	0.0048	0.0518	0.0919	10.8799	
Herz	0.0127	0.0275	0.4626	2.1618	15
Hoden	0.0115	0.0117	0.9839	1.0163	
Lunge	0.0114	0.0061	1.8628	0.5368	
Magen-Speiserohre	0.0000	0.0460	0.0000	undef	
Muskel-Skelett	0.0154	0.0060	2.5700	0.3891	
Niere	0.0054	0.0068	0.7930	1.2610	20
Pankreas	0.0050	0.0331	0.1496	6.6857	
Penis	0.0090	0.0533	0.1685	5.9357	
Prostata	0.0174	0.0192	0.9099	1.0990	
Uterus_Endometrium	0.0068	0.0000	undef	0.0000	
Uterus_Myometrium	0.0152	0.0000	undef	0.0000	25
Uterus_allgemein	0.0204	0.0000	undef	0.0000	
Brust-Hyperplasie	0.0064				
Prostata-Hyperplasie	0.0238				
Samenblase	0.0000				
Sinnesorgane	0.0000				30
Weisse_Blutkoerperchen	0.0251				
Zervix	0.0106				
					35
					40
					45
					50
					55
					60
					65

## FOETUS

%Haeufigkeit

Entwicklung	0.0000
Gastrointestinal	0.0167
Gehirn	0.0438
Haematopoetisch	0.0118
Haut	0.0000
Hepatisch	0.0000
Herz-Blutgefuesse	0.0107
Lunge	0.0181
Nebenniere	0.0000
Niere	0.0247
Placenta	0.0061
Prostata	0.0249
Sinnesorgane	0.0000

## NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN

%Haeufigkeit

Brust	0.0408
Eierstock_n	0.0000
Eierstock_t	0.0101
Endokrines_Gewebe	0.0000
Foetal	0.0087
Gastrointestinal	0.0122
Haematopoetisch	0.0000
Haut-Muskel	0.0130
Hoden	0.0000
Lunge	0.0082
Nerven	0.0010
Prostata	0.0068
Sinnesorgane	0.0000
Uterus_n	0.0000



## DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 18

	NORMAL	TUMOR	Verhaeltnisse	
	%Haeufigkeit	%Haeufigkeit	N/T	T/N
5	Blase 0.0195	0.2556	0.0763	13.1109
	Brust 0.0166	0.0357	0.4656	2.1477
	Duendarm 0.0061	0.0662	0.0927	10.7893
	Eierstock 0.0389	0.0052	7.4832	0.1336
	Endokrines_Gewebe 0.0392	0.0326	1.2017	0.8321
10	Gastrointestinal 0.0019	0.0000	undef	0.0000
	Gehirn 0.0007	0.0606	0.0122	81.9491
	Haematopoetisch 0.0107	0.0000	undef	0.0000
	Haut 0.0220	0.5085	0.0433	23.0839
	Hepatisch 0.0238	0.0518	0.4596	2.1760
15	Herz 0.0085	0.0000	undef	0.0000
	Hoden 0.0115	0.0000	undef	0.0000
	Lunge 0.0104	0.0041	2.5402	0.3937
	Magen-Speiserohre 0.0000	0.0077	0.0000	undef
	Muskel-Skelett 0.0600	0.0480	1.2493	0.8005
20	Niere 0.0407	0.0068	5.9478	0.1681
	Pankreas 0.0198	0.0331	0.5983	1.6714
	Penis 0.0030	0.1066	0.0281	35.6140
	Prostata 0.0000	0.0021	0.0000	undef
	Uterus_Endometrium 0.0405	0.0000	undef	0.0000
	Uterus_Myometrium 0.0305	0.1155	0.2641	3.7870
25	Uterus_allgemein 0.0153	0.0000	undef	0.0000
	Brust-Hyperplasie 0.0064			
	Prostata-Hyperplasie 0.0030			
	Samenblase 0.0000			
	Sinnesorgane 0.0118			
30	Weisse_Blutkoerperchen 0.0000			
	Zervix 0.0000			

	FOETUS
	%Haeufigkeit
35	Entwicklung 0.0696
	Gastrointestinal 0.3332
	Gehirn 0.0000
	Haematopoetisch 0.2202
40	Haut 0.0000
	Hepatisch 1.6381
	Herz-Blutgefuesse 0.0285
	Lunge 0.1337
	Nebenniere 1.0903
45	Niere 0.6301
	Placenta 0.6786
	Prostata 0.0499
	Sinnesorgane 0.0000

	NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN
	%Haeufigkeit
50	Brust 0.0544
	Eierstock_n 0.0000
55	Eierstock_t 0.1063
	Endokrines_Gewebe 0.0000
	Foetal 0.4264
	Gastrointestinal 0.0000
	Haematopoetisch 0.0000
60	Haut-Muskel 0.0000
	Hoden 0.0000
	Lunge 0.0000
	Nerven 0.0030
	Prostata 0.0000
65	Sinnesorgane 0.0000
	Uterus_n 0.0250

## DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 19

	NORMAL	TUMOR	Verhaeltnisse		
	%Haeufigkeit	%Haeufigkeit	N/T	T/N	
Blase	0.0195	0.2301	0.0847	11.7998	5
Brust	0.0192	0.0113	1.7013	0.5878	
Duenn darm	0.0061	0.0331	0.1854	5.3946	
Eierstock	0.0180	0.0000	undef	0.0000	
Endokrines Gewebe	0.0034	0.1555	0.0219	45.6387	
Gastrointestinal	0.0000	0.0370	0.0000	undef	10
Gehirn	0.0214	0.1561	0.1374	7.2801	
Haematopoetisch	0.0134	0.0000	undef	0.0000	
Haut	0.0073	0.0000	undef	0.0000	
Hepatisch	0.0048	0.0000	undef	0.0000	
Herz	0.0085	0.0000	undef	0.0000	15
Hoden	0.0000	0.0000	undef	undef	
Lunge	0.0197	0.0020	9.6527	0.1036	
Magen-Speiserohre	0.0000	0.0077	0.0000	undef	
Muskel-Skelett	0.0034	0.1320	0.0260	38.5221	
Niere	0.0109	0.0068	1.5861	0.6305	20
Pankreas	0.0083	0.0000	undef	0.0000	
Penis	0.0000	0.0000	undef	undef	
Prostata	0.0000	0.0021	0.0000	undef	
Uterus Endometrium	0.0203	0.0000	undef	0.0000	
Uterus Myometrium	0.0000	0.0475	0.0000	undef	25
Uterus allgemein	0.0153	0.0000	undef	0.0000	
Brust-Hyperplasie	0.0128				
Prostata-Hyperplasie	0.0030				
Samenblase	0.0000				
Sinnesorgane	0.0235				30
Weisse Blutkoerperchen	0.0000				
Zervix	0.0213				
					35
Entwicklung	0.0417				
Gastrointestinal	0.1361				
Gehirn	0.0063				
Haematopoetisch	0.1337				
Haut	0.0000				40
Hepatisch	0.3380				
Herz-Blutgefuesse	0.0249				
Lunge	0.0578				
Nebenniere	0.5071				
Niere	0.2594				45
Placenta	0.4120				
Prostata	0.0000				
Sinnesorgane	0.0000				
					50
					55
					60
					65

## NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN

%Haeufigkeit

Brust	0.0612	
Eierstock_n	0.0000	
Eierstock_t	0.0000	
Endokrines Gewebe	0.0000	55
Foetal	0.1188	
Gastrointestinal	0.0244	
Haematopoetisch	0.0000	
Haut-Muskel	0.0000	
Hoden	0.0154	60
Lunge	0.0000	
Nerven	0.0141	
Prostata	0.0000	
Sinnesorgane	0.0000	
Uterus_n	0.0416	65

# DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 20

		NORMAL	TUMOR	Verhaeltnisse	
		%Haeufigkeit	%Haeufigkeit	N/T	T/N
5	Blase	0.0039	0.0383	0.1017	9.8332
	Brust	0.0077	0.0075	1.0208	0.9796
	Duennndarm	0.0061	0.0496	0.1236	8.0920
	Eierstock	0.0000	0.0130	0.0000	undef
	Endokrines_Gewebe	0.0153	0.0000	undef	0.0000
10	Gastrointestinal	0.0115	0.0000	undef	0.0000
	Gehirn	0.0022	0.0031	0.7200	1.3890
	Haematopoetisch	0.0013	0.0000	undef	0.0000
	Haut	0.0073	0.0000	undef	0.0000
	Hepatisch	0.0048	0.0129	0.3676	2.7200
15	Herz	0.0000	0.0000	undef	undef
	Hoden	0.0000	0.0000	undef	undef
	Lunge	0.0042	0.0102	0.4064	2.4605
	Magen-Speiseroehre	0.0290	0.0077	3.7816	0.2644
	Muskel-Skelett	0.0000	0.0000	undef	undef
20	Niere	0.0136	0.0000	undef	0.0000
	Pankreas	0.0017	0.0000	undef	0.0000
	Penis	0.0030	0.0000	undef	0.0000
	Prostata	0.0065	0.0106	0.6142	1.6282
	Uterus_Endometrium	0.0068	0.0528	0.1280	7.8106
25	Uterus_Myometrium	0.0000	0.0000	undef	undef
	Uterus_allgemein	0.0000	0.0000	undef	undef
	Brust-Hyperplasie	0.0032			
	Prostata-Hyperplasie	0.0268			
	Samenblase	0.0267			
	Sinnesorgane	0.0235			
30	Weisse_Blutkoerperchen	0.0000			
	Zervix	0.0319			

		FOETUS
		%Haeufigkeit
35	Entwicklung	0.0139
	Gastrointestinal	0.0139
	Gehirn	0.0000
	Haematopoetisch	0.0039
40	Haut	0.0000
	Hepatisch	0.0000
	Herz-Blutgefuesse	0.0000
	Lunge	0.0036
	Nebenniere	0.0254
45	Niere	0.0062
	Placenta	0.0000
	Prostata	0.0000
	Sinnesorgane	0.0000

		NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN
		%Haeufigkeit
50	Brust	0.0204
	Eierstock_n	0.0000
55	Eierstock_t	0.0203
	Endokrines_Gewebe	0.0245
	Foetal	0.0128
	Gastrointestinal	0.0122
	Haematopoetisch	0.0000
60	Haut-Muskel	0.0000
	Hoden	0.0154
	Lunge	0.0082
	Nerven	0.0090
	Prostata	0.0068
65	Sinnesorgane	0.0000
	Uterus_n	0.0000

## DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 21

	NORMAL	TUMOR	Verhaeltnisse		
	%Haeufigkeit	%Haeufigkeit	N/T	T/N	
Blase	0.0039	0.0358	0.1090	9.1777	5
Brust	0.0077	0.0207	0.3712	2.6940	
Duenn darm	0.0215	0.0331	0.6488	1.5413	
Eierstock	0.0030	0.0078	0.3838	2.6058	
Endokrines Gewebe	0.0000	0.0000	undef	undef	10
Gastrointestinal	0.0575	0.0046	12.4251	0.0805	
Gehirn	0.0000	0.0072	0.0000	undef	
Haematopoetisch	0.0214	0.0000	undef	0.0000	
Haut	0.0037	0.0000	undef	0.0000	15
Hepatisch	0.0000	0.0194	0.0000	undef	
Herz	0.0000	0.0137	0.0000	undef	
Hoden	0.0000	0.0585	0.0000	undef	
Lunge	0.0104	0.0225	0.4618	2.1652	20
Magen-Speiserohre	0.0000	0.0000	undef	undef	
Muskel-Skelett	0.0257	0.0120	2.1416	0.4669	
Niere	0.0054	0.0000	undef	0.0000	
Pankreas	0.0099	0.0221	0.4487	2.2286	25
Penis	0.0000	0.0000	undef	undef	
Prostata	0.0022	0.0000	undef	0.0000	
Uterus_Endometrium	0.0000	0.0000	undef	undef	
Uterus_Myometrium	0.0000	0.0000	undef	undef	30
Uterus_allgemein	0.0000	0.0000	undef	undef	
Brust-Hyperplasie	0.0096				
Prostata-Hyperplasie	0.0000				
Samenblase	0.0000				35
Sinnesorgane	0.0000				
Weisse_Blutkoerperchen	0.0000				
Zervix	0.0000				

FOETUS					
	%Haeufigkeit				
Entwicklung	0.0000				35
Gastrointestinal	0.0000				
Gehirn	0.0000				
Haematopoetisch	0.0000				
Haut	0.0000				40
Hepatisch	0.0000				
Herz-Blutgefuesse	0.0000				
Lunge	0.0000				
Nebenniere	0.0000				45
Niere	0.0000				
Placenta	0.0000				
Prostata	0.0000				
Sinnesorgane	0.0000				50

NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN					
	%Haeufigkeit				
Brust	0.0136				55
Eierstock_n	0.0000				
Eierstock_t	0.0051				
Endokrines_Gewebe	0.0000				
Foetal	0.0000				60
Gastrointestinal	0.0122				
Haematopoetisch	0.0000				
Haut-Muskel	0.0000				
Hoden	0.0000				65
Lunge	0.0164				
Nerven	0.0000				
Prostata	0.0000				
Sinnesorgane	0.0000				65
Uterus_n	0.0000				

## DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 22

	NORMAL	TUMOR	Verhaeltnisse	
	%Haefufigkeit	%Haefufigkeit	N/T	T/N
5	Blase 0.0351	0.3144	0.1116	8.9591
	Brust 0.0281	0.0470	0.5989	1.6698
	Duennndarm 0.0092	0.0662	0.1390	7.1929
	Eierstock 0.0569	0.0208	2.7342	0.3657
	Endokrines_Gewebe 0.0596	0.0527	1.1321	0.8833
10	Gastrointestinal 0.0019	0.0139	0.1381	7.2434
	Gehirn 0.0022	0.0945	0.0235	42.5950
	Haematopoetisch 0.0174	0.0000	undef	0.0000
	Haut 0.0220	0.5085	0.0433	23.0839
	Hepatisch 0.0285	0.0582	0.4902	2.0400
15	Herz 0.0223	0.0000	undef	0.0000
	Hoden 0.0173	0.0000	undef	0.0000
	Lunge 0.0145	0.0082	1.7781	0.5624
	Magen-Speiserohre 0.0290	0.0077	3.7816	0.2644
	Muskel-Skelett 0.0788	0.0540	1.4595	0.6852
20	Niere 0.0489	0.0137	3.5687	0.2802
	Pankreas 0.0264	0.0442	0.5983	1.6714
	Penis 0.0090	0.1066	0.0842	11.8713
	Prostata 0.0000	0.0064	0.0000	undef
	Uterus_Endometrium 0.1013	0.0000	undef	0.0000
	Uterus_Myometrium 0.0381	0.1494	0.2551	3.9206
25	Uterus_allgemein 0.0153	0.0954	0.1601	6.2452
	Brust-Hyperplasie 0.0096			
	Prostata-Hyperplasie 0.0059			
	Samenblase 0.0000			
	Sinnesorgane 0.0118			
30	Weisse_Blutkoerperchen 0.0000			
	Zervix 0.0000			

	FOETUS
	%Haefufigkeit
35	Entwicklung 0.0696
	Gastrointestinal 0.4554
	Gehirn 0.0000
	Haematopoetisch 0.2753
40	Haut 0.0000
	Hepatisch 1.6381
	Herz-Blutgefuesse 0.0605
	Lunge 0.1879
	Nebenniere 1.1663
45	Niere 0.8215
	Placenta 0.7816
	Prostata 0.0499
	Sinnesorgane 0.0000

	NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN
	%Haefufigkeit
50	Brust 0.0612
	Eierstock_n 0.0000
	Eierstock_t 0.1164
55	Endokrines_Gewebe 0.0000
	Foetal 0.4665
	Gastrointestinal 0.0000
	Haematopoetisch 0.0000
60	Haut-Muskel 0.0000
	Hoden 0.0000
	Lunge 0.0000
	Nerven 0.0030
	Prostata 0.0068
65	Sinnesorgane 0.0000
	Uterus_n 0.0291

## DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 23

	NORMAL	TUMOR	Verhaeltnisse		
	%Haeufigkeit	%Haeufigkeit	N/T	T/N	
Blase	0.0039	0.0332	0.1173	8.5221	5
Brust	0.0038	0.0263	0.1458	6.8574	
Duennndarm	0.0184	0.0331	0.5561	1.7982	
Eierstock	0.0150	0.0156	0.9594	1.0423	
Endokrines_Gewebe	0.0170	0.0075	2.2642	0.4417	
Gastrointestinal	0.0192	0.0324	0.5917	1.6901	10
Gehirn	0.0133	0.0113	1.1781	0.8488	
Haematopoetisch	0.0174	0.0000	undef	0.0000	
Haut	0.0220	0.0000	undef	0.0000	
Hepatisch	0.0048	0.0388	0.1225	8.1599	
Herz	0.0339	0.0137	2.4671	0.4053	15
Hoden	0.0288	0.0234	1.2299	0.8130	
Lunge	0.0218	0.0184	1.1854	0.8436	
Magen-Speiserohre	0.0000	0.0307	0.0000	undef	
Muskel-Skelett	0.0291	0.0240	1.2136	0.8240	
Niere	0.0136	0.0137	0.9913	1.0088	20
Pankreas	0.0050	0.0166	0.2991	3.3428	
Penis	0.0120	0.0000	undef	0.0000	
Prostata	0.0153	0.0149	1.0236	0.9769	
Uterus_Endometrium	0.0203	0.0000	undef	0.0000	
Uterus_Myometrium	0.0000	0.0204	0.0000	undef	25
Uterus_allgemein	0.0051	0.0000	undef	0.0000	
Brust-Hyperplasie	0.0128				
Prostata-Hyperplasie	0.0178				
Samenblase	0.0267				
Sinnesorgane	0.0118				30
Weisse_Blutkoerperchen	0.0243				
Zervix	0.0213				

	FOETUS	
	%Haeufigkeit	
Entwicklung	0.0139	35
Gastrointestinal	0.0056	
Gehirn	0.0125	
Haematopoetisch	0.0000	
Haut	0.0000	
Hepatisch	0.0000	40
Herz-Blutgefuesse	0.0320	
Lunge	0.0036	
Nebenniere	0.0254	
Niere	0.0124	
Placenta	0.0121	45
Prostata	0.0000	
Sinnesorgane	0.0000	

	NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN	
	%Haeufigkeit	
Brust	0.0204	50
Eierstock_n	0.0000	
Eierstock_t	0.0354	
Endokrines_Gewebe	0.0000	
Foetal	0.0105	
Gastrointestinal	0.0000	55
Haematopoetisch	0.0171	
Haut-Muskel	0.0454	
Hoden	0.0000	
Lunge	0.0246	
Nerven	0.0211	60
Prostata	0.0000	
Sinnesorgane	0.0000	
Uterus_n	0.0000	65

## DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 24

		NORMAL	TUMOR	Verhaeltnisse	
		%Haeufigkeit	%Haeufigkeit	N/T	T/N
5	Blase	0.0195	0.1457	0.1338	7.4732
	Brust	0.0013	0.0169	0.0756	13.2250
	Duennndarm	0.0061	0.0496	0.1236	8.0920
	Eierstock	0.0090	0.0260	0.3454	2.8954
	Endokrines_Gewebe	0.0119	0.0125	0.9509	1.0516
10	Gastrointestinal	0.1648	0.0231	7.1237	0.1404
	Gehirn	0.0030	0.0308	0.0960	10.4173
	Haematopoetisch	0.2099	0.0000	undef	0.0000
	Haut	0.0000	0.0000	undef	undef
	Hepatisch	0.0000	0.2006	0.0000	undef
15	Herz	0.0032	0.0137	0.2313	4.3235
	Hoden	0.0000	0.0819	0.0000	undef
	Lunge	0.0956	0.2965	0.3223	3.1023
	Magen-Speiserohre	0.0097	0.1917	0.0504	19.8329
	Muskel-Skelett	0.0959	0.0240	3.9977	0.2501
20	Niere	0.0081	0.0205	0.3965	2.5219
	Pankreas	0.0182	0.0497	0.3656	2.7350
	Penis	0.0030	0.1333	0.0225	44.5175
	Prostata	0.0131	0.0064	2.0473	0.4885
	Uterus_Endometrium	0.0000	0.0000	undef	undef
25	Uterus_Myometrium	0.0000	0.0000	undef	undef
	Uterus_allgemein	0.0051	0.0000	undef	0.0000
	Brust-Hyperplasie	0.0096			
	Prostata-Hyperplasie	0.0030			
	Samenblase	0.0267			
	Sinnesorgane	0.0706			
30	Weisse_Blutkoerperchen	0.0017			
	Zervix	0.0000			
35		FOETUS			
		%Haeufigkeit			
	Entwicklung	0.0000			
	Gastrointestinal	0.0000			
	Gehirn	0.0000			
	Haematopoetisch	0.0000			
40	Haut	0.0000			
	Hepatisch	0.0000			
	Herz-Blutgefuesse	0.0000			
	Lunge	0.0000			
	Nebenniere	0.0000			
45	Niere	0.0000			
	Placenta	0.0000			
	Prostata	0.0000			
	Sinnesorgane	0.0000			
50					
		NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN			
		%Haeufigkeit			
	Brust	0.0068			
	Eierstock_n	0.0000			
55	Eierstock_t	0.2076			
	Endokrines_Gewebe	0.0000			
	Foetal	0.0006			
	Gastrointestinal	0.0000			
	Haematopoetisch	0.0000			
60	Haut-Muskel	0.0000			
	Hoden	0.0000			
	Lunge	0.0082			
	Nerven	0.0030			
	Prostata	0.0068			
65	Sinnesorgane	0.0000			
	Uterus_n	0.0042			

# DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 25

	NORMAL	TUMOR	Verhaeltnisse	
	%Haeufigkeit	%Haeufigkeit	N/T T/N	
Blase	0.0039	0.0281	0.1387 7.2110	5
Brust	0.0217	0.0132	1.6527 0.6051	
Duennndarm	0.0123	0.0000	undef 0.0000	
Eierstock	0.0000	0.0026	0.0000 undef	
Endokrines_Gewebe	0.0102	0.0150	0.6792 1.4722	
Gastrointestinal	0.0115	0.0046	2.4850 0.4024	10
Gehirn	0.0148	0.0154	0.9599 1.0417	
Haematopoetisch	0.0107	0.0000	undef 0.0000	
Haut	0.0220	0.0847	0.2599 3.8473	
Hepatisch	0.0000	0.0000	undef undef	
Herz	0.0085	0.0000	undef 0.0000	15
Hoden	0.0058	0.0117	0.4920 2.0326	
Lunge	0.0145	0.0061	2.3708 0.4218	
Magen-Speiseroehre	0.0000	0.0077	0.0000 undef	
Muskel-Skelett	0.0086	0.0360	0.2380 4.2024	
Niere	0.0081	0.0000	undef 0.0000	20
Pankreas	0.0050	0.0166	0.2991 3.3428	
Penis	0.0030	0.0000	undef 0.0000	
Prostata	0.0109	0.0064	1.7060 0.5862	
Uterus_Endometrium	0.0203	0.0000	undef 0.0000	
Uterus_Myometrium	0.0076	0.0068	1.1223 0.8911	25
Uterus_allgemein	0.0102	0.0000	undef 0.0000	
Brust-Hyperplasie	0.0096			
Prostata-Hyperplasie	0.0059			
Samenblase	0.0089			
Sinnesorgane	0.0000			
Weisse_Blutkoerperchen	0.0286			30
Zervix	0.0106			
	FOETUS			
	%Haeufigkeit			35
Entwicklung	0.0000			
Gastrointestinal	0.0028			
Gehirn	0.0125			
Haematopoetisch	0.0000			
Haut	0.0000			40
Hepatisch	0.0000			
Herz-Blutgefuesse	0.0000			
Lunge	0.0036			
Nebenniere	0.0000			
Niere	0.0000			45
Placenta	0.0061			
Prostata	0.0000			
Sinnesorgane	0.0000			
	NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN			50
	%Haeufigkeit			
Brust	0.0204			
Eierstock_n	0.0000			
Eierstock_t	0.0051			
Endokrines_Gewebe	0.0000			55
Foetal	0.0029			
Gastrointestinal	0.0000			
Haematopoetisch	0.0000			
Haut-Muskel	0.0000			
Hoden	0.0000			60
Lunge	0.0000			
Nerven	0.0010			
Prostata	0.0068			
Sinnesorgane	0.0000			
Uterus_n	0.0000			65



## DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 26

		NORMAL	TUMOR	Verhaeltnisse	
		%Haeufigkeit	%Haeufigkeit	N/T	T/N
5	Blase	0.0039	0.0256	0.1525	6.5555
	Brust	0.0038	0.0056	0.6805	1.4694
	Duennndarm	0.0031	0.0000	undef	0.0000
	Eierstock	0.0210	0.0052	4.0294	0.2482
	Endokrines_Gewebe	0.0017	0.0000	undef	0.0000
10	Gastrointestinal	0.0057	0.0000	undef	0.0000
	Gehirn	0.0037	0.0031	1.1999	0.8334
	Haematopoetisch	0.0040	0.0000	undef	0.0000
	Haut	0.0073	0.0000	undef	0.0000
	Hepatisch	0.0000	0.0000	undef	undef
15	Herz	0.0000	0.0000	undef	undef
	Hoden	0.0000	0.0000	undef	undef
	Lunge	0.0073	0.0020	3.5562	0.2812
	Magen-Speiserohre	0.0000	0.0000	undef	undef
	Muskel-Skelett	0.0069	0.0000	undef	0.0000
	Niere	0.0027	0.0000	undef	0.0000
20	Pankreas	0.0017	0.0000	undef	0.0000
	Penis	0.0030	0.0000	undef	0.0000
	Prostata	0.0000	0.0021	0.0000	undef
	Uterus_Endometrium	0.0000	0.0000	undef	undef
	Uterus_Myometrium	0.0076	0.0068	1.1223	0.8911
25	Uterus_allgemein	0.0102	0.0000	undef	0.0000
	Brust-Hyperplasie	0.0000			
	Prostata-Hyperplasie	0.0030			
	Samenblase	0.0089			
	Sinnesorgane	0.0000			
30	Weisse_Blutkoerperchen	0.0121			
	Zervix	0.0000			

		FOETUS
		%Haeufigkeit
35	Entwicklung	0.0000
	Gastrointestinal	0.0139
	Gehirn	0.0063
	Haematopoetisch	0.0039
40	Haut	0.0000
	Hepatisch	0.0000
	Herz-Blutgefasse	0.0036
	Lunge	0.0036
	Nebenniere	0.0000
45	Niere	0.0062
	Placenta	0.0000
	Prostata	0.0000
	Sinnesorgane	0.0000,

		NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN
		%Haeufigkeit
50	Brust	0.0000
	Eierstock_n	0.0000
	Eierstock_t	0.0101
55	Endokrines_Gewebe	0.0000
	Foetal	0.0029
	Gastrointestinal	0.0488
	Haematopoetisch	0.0114
60	-- Haut-Muskel	0.0130
	Hoden	0.0154
	Lunge	0.0000
	Nerven	0.0020
	Prostata	0.0274
	Sinnesorgane	0.0155
65	Uterus_n	0.0083

# DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 27

	NORMAL	TUMOR	Verhaeltnisse		
	%Haeufigkeit	%Haeufigkeit	N/T	T/N	
Blase	0.0039	0.0256	0.1525	6.5555	5
Brust	0.0115	0.0150	0.7656	1.3062	
Duennndarm	0.0307	0.0000	undef	0.0000	
Eierstock	0.0150	0.0052	2.8781	0.3474	
Endokrines_Gewebe	0.0119	0.0150	0.7925	1.2619	
Gastrointestinal	0.0057	0.0139	0.4142	2.4145	10
Gehirn	0.0074	0.0092	0.8000	1.2501	
Haematopoetisch	0.0147	0.0000	undef	0.0000	
Haut	0.0037	0.1695	0.0217	46.1678	
Hepatisch	0.0048	0.0000	undef	0.0000	
Herz	0.0074	0.0137	0.5397	1.8529	15
Hoden	0.0000	0.0234	0.0000	undef	
Lunge	0.0104	0.0082	1.2701	0.7873	
Magen-Speiseroehre	0.0000	0.0230	0.0000	undef	
Muskel-Skelett	0.0086	0.0000	undef	0.0000	
Niere	0.0190	0.0000	undef	0.0000	20
Pankreas	0.0066	0.0055	1.1966	0.8357	
Penis	0.0150	0.0267	0.5616	1.7807	
Prostata	0.0087	0.0149	0.5849	1.7096	
Uterus_Endometrium	0.0203	0.0000	undef	0.0000	
Uterus_Myometrium	0.0000	0.0136	0.0000	undef	25
Uterus_allgemein	0.0051	0.0000	undef	0.0000	
Brust-Hyperplasie	0.0064				
Prostata-Hyperplasie	0.0178				
Samenblase	0.0178				
Sinnesorgane	0.0235				
Weisse_Blutkoerperchen	0.0104				30
Zervix	0.0000				
FOETUS					
	%Haeufigkeit				35
Entwicklung	0.0000				
Gastrointestinal	0.0111				
Gehirn	0.0125				
Haematopoetisch	0.0000				
Haut	0.0000				40
Hepatisch	0.0260				
Herz-Blutgefuesse	0.0071				
Lunge	0.0217				
Nebenniere	0.0254				
Niere	0.0062				45
Placenta	0.0061				
Prostata	0.0000				
Sinnesorgane	0.0000				
NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN					
	%Haeufigkeit				50
Brust	0.0136				
Eierstock_n	0.0000				
Eierstock_t	0.0101				
Endokrines_Gewebe	0.0245				55
Foetal	0.0181				
Gastrointestinal	0.0244				
Haematopoetisch	0.0114				
Haut-Muskel	0.0065				60
Hoden	0.0000				
Lunge	0.0164				
Nerven	0.0090				
Prostata	0.0068				
Sinnesorgane	0.0232				
Uterus_n	0.0000				65

		NORMAL	TUMOR	Verhaeltnisse	
		%Haeufigkeit	%Haeufigkeit	N/T	T/N
5	Blase	0.0312	0.1968	0.1585	6.3096
	Brust	0.0281	0.0526	0.5347	1.8702
	Duenn darm	0.2177	0.0000	undef	0.0000
	Eierstock	0.0000	0.0494	0.0000	undef
	Endokrines_Gewebe	0.0017	0.0000	undef	0.0000
10	Gastrointestinal	0.2778	0.0324	8.5792	0.1166
	Gehirn	0.0007	0.0133	0.0554	18.0566
	Haematopoetisch	0.1470	0.0000	undef	0.0000
	Haut	0.0551	0.0000	undef	0.0000
	Hepatisch	0.0000	0.1876	0.0000	undef
15	Herz	0.0011	0.0137	0.0771	12.9706
	Hoden	0.0000	0.2339	0.0000	undef
	Lunge	0.0623	0.1206	0.5166	1.9356
	Magen-Speiserohre	0.0000	0.0153	0.0000	undef
	Muskel-Skelett	0.1216	0.0300	4.0548	0.2466
20	Niere	0.0081	0.1027	0.0793	12.6097
	Pankreas	0.0743	0.2209	0.3365	2.9714
	Penis	0.0000	0.0000	undef	undef
	Prostata	0.0065	0.0149	0.4387	2.2795
	Uterus_Endometrium	0.0068	0.0000	undef	0.0000
	Uterus_Myometrium	0.0000	0.0000	undef	undef
25	Uterus_allgemein	0.0000	0.0000	undef	undef
	Brust-Hyperplasie	0.0384			
	Prostata-Hyperplasie	0.0030			
	Samenblase	0.0000			
	Sinnesorgane	0.0118			
30	Weisse_Blutkoerperchen	0.0017			
	Zervix	0.0000			

		FOETUS
		%Haeufigkeit
35	Entwicklung	0.0000
	Gastrointestinal	0.0000
	Gehirn	0.0000
	Haematopoetisch	0.0079
40	Haut	0.0000
	Hepatisch	0.0000
	Herz-Blutgefuesse	0.0000
	Lunge	0.0000
	Nebenniere	0.0000
45	Niere	0.0000
	Placenta	0.0000
	Prostata	0.0000
	Sinnesorgane	0.0000

		NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN
		%Haeufigkeit
50	Brust	0.0068
	Eierstock_n	0.0000
	Eierstock_t	0.0051
55	Endokrines_Gewebe	0.0000
	Foetal	0.0000
	Gastrointestinal	0.0000
	Haematopoetisch	0.0000
60	Haut-Muskel	0.0000
	Hoden	0.0000
	Lunge	0.0082
	Nerven	0.0000
	Prostata	0.0000
	Sinnesorgane	0.0000
65	Uterus_n	0.0000

# DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 29

	NORMAL	TUMOR	Verhaeltnisse	
	%Haeufigkeit	%Haeufigkeit	N/T T/N	
Blase	0.0273	0.1713	0.1594 6.2745	5
Brust	0.0217	0.0395	0.5509 1.8152	
Duenn darm	0.1901	0.0000	undef 0.0000	
Eierstock	0.0000	0.0390	0.0000 undef	
Endokrines Gewebe	0.0017	0.0000	undef 0.0000	
Gastrointestinal	0.2491	0.0278	8.9737 0.1114	10
Gehirn	0.0000	0.0113	0.0000 undef	
Haematopoetisch	0.1297	0.0000	undef 0.0000	
Haut	0.0477	0.0000	undef 0.0000	
Hepatisch	0.0000	0.1423	0.0000 undef	
Herz	0.0032	0.0000	undef 0.0000	
Hoden	0.0000	0.2105	0.0000 undef	15
Lunge	0.0457	0.0818	0.5588 1.7894	
Magen-Speiserohre	0.0000	0.0153	0.0000 undef	
Muskel-Skelett	0.0822	0.0240	3.4266 0.2918	
Niere	0.0054	0.1027	0.0529 18.9146	20
Pankreas	0.0694	0.2430	0.2855 3.5020	
Penis	0.0000	0.0000	undef undef	
Prostata	0.0153	0.0106	1.4331 0.6978	
Uterus_Endometrium	0.0068	0.0000	undef 0.0000	
Uterus_Myometrium	0.0000	0.0000	undef undef	
Uterus_allgemein	0.0000	0.0000	undef undef	25
Brust-Hyperplasie	0.0288			
Prostata-Hyperplasie	0.0030			
Samenblase	0.0000			
Sinnesorgane	0.0118			
Weisse_Blutkoerperchen	0.0009			30
Zervix	0.0000			

	FOETUS	
	%Haeufigkeit	
Entwicklung	0.0000	35
Gastrointestinal	0.0000	
Gehirn	0.0000	
Haematopoetisch	0.0079	
Haut	0.0000	40
Hepatisch	0.0000	
Herz-Blutgefuesse	0.0000	
Lunge	0.0000	
Nebenniere	0.0000	
Niere	0.0000	45
Placenta	0.0000	
Prostata	0.0000	
Sinnesorgane	0.0000	

	NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN	
	%Haeufigkeit	
Brust	0.0068	
Eierstock_n	0.0000	
Eierstock_t	0.0000	
Endokrines_Gewebe	0.0000	55
Foetal	0.0000	
Gastrointestinal	0.0000	
Haematopoetisch	0.0000	
Haut-Muskel	0.0000	
Hoden	0.0000	60
Lunge	0.0082	
Nerven	0.0000	
Prostata	0.0000	
Sinnesorgane	0.0000	
Uterus_n	0.0000	65

## DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 30

		NORMAL	TUMOR	Verhaeltnisse	
		%Haeufigkeit	%Haeufigkeit	N/T	T/N
5	Blase	0.0858	0.4525	0.1896	5.2742
	Brust	0.1279	0.0977	1.3087	0.7641
	Duennndarm	0.4998	0.2150	2.3242	0.4302
	Eierstock	0.0180	0.1353	0.1328	7.5280
	Endokrines Gewebe	0.0307	0.0176	1.7466	0.5725
10	Gastrointestinal	0.7434	0.1341	5.5413	0.1805
	Gehirn	0.0030	0.0524	0.0565	17.7093
	Haematopoetisch	0.6884	0.0000	undef	0.0000
	Haut	0.0551	0.1695	0.3249	3.0779
	Hepatisch	0.0381	0.7635	0.0499	20.0598
15	Herz	0.0095	0.0412	0.2313	4.3235
	Hoden	0.0000	0.1403	0.0000	undef
	Lunge	0.2773	0.5418	0.5119	1.9536
	Magen-Speiserohre	0.0676	0.3450	0.1961	5.0999
	Muskel-Skelett	0.3203	0.0660	4.8544	0.2060
20	Niere	0.0163	0.1780	0.0915	10.9284
	Pankreas	0.0529	0.2319	0.2279	4.3875
	Penis	0.0000	0.0000	undef	undef
	Prostata	0.0828	0.0192	4.3220	0.2314
	Uterus_Endometrium	0.0068	0.0000	undef	0.0000
	Uterus_Myometrium	0.0076	0.0204	0.3741	2.6732
25	Uterus_allgemein	0.0102	0.0000	undef	0.0000
	Brust-Hyperplasie	0.1343			
	Prostata-Hyperplasie	0.0268			
	Samenblase	0.0267			
	Sinnesorgane	0.0353			
30	Weisse_Blutkoerperchen	0.0043			
	Zervix	0.0852			

		FOETUS
		%Haeufigkeit
35	Entwicklung	0.0000
	Gastrointestinal	0.0000
	Gehirn	0.0000
	Haematopoetisch	0.0118
40	Haut	0.0000
	Hepatisch	0.0000
	Herz-Blutgefuesse	0.0000
	Lunge	0.0000
	Nebenniere	0.0000
45	Niere	0.0000
	Placenta	0.0000
	Prostata	0.0000
	Sinnesorgane	0.0000

		NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN
		%Haeufigkeit
50	Brust	0.1837
	Eierstock_n	0.0000
	Eierstock_t	0.0405
55	Endokrines Gewebe	0.0000
	Foetal	0.0023
	Gastrointestinal	0.0976
	Haematopoetisch	0.0000
60	Haut-Muskel	0.0000
	Hoden	0.0000
	Lunge	0.0000
	Nerven	0.0020
	Prostata	0.0000
65	Sinnesorgane	0.0000
	Uterus_n	0.0042

# DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 31

	NORMAL %Haeufigkeit	TUMOR %Haeufigkeit	Verhaeltnisse		
			N/T	T/N	
Blase	0.0039	0.0204	0.1907	5.2444	5
Brust	0.0077	0.0056	1.3611	0.7347	
Duenn darm	0.0184	0.0331	0.5561	1.7982	
Eierstock	0.0150	0.0104	1.4391	0.6949	
Endokrines Gewebe	0.0102	0.0000	undef	0.0000	
Gastrointestinal	0.0019	0.0046	0.4142	2.4145	10
Gehirn	0.0074	0.0072	1.0285	0.9723	
Haematopoetisch	0.0067	0.0000	undef	0.0000	
Haut	0.0147	0.0000	undef	0.0000	
Hepatisch	0.0000	0.0065	0.0000	undef	
Herz	0.0064	0.0000	undef	0.0000	15
Hoden	0.0345	0.0234	1.4759	0.6775	
Lunge	0.0073	0.0143	0.5080	1.9684	
Magen-Speiserohre	0.0000	0.0000	undef	undef	
Muskel-Skelett	0.0034	0.0000	undef	0.0000	
Niere	0.0054	0.0000	undef	0.0000	20
Pankreas	0.0050	0.0055	0.8974	1.1143	
Penis	0.0030	0.0000	undef	0.0000	
Prostata	0.0153	0.0043	3.5827	0.2791	
Uterus_Endometrium	0.0068	0.0000	undef	0.0000	
Uterus_Myometrium	0.0000	0.0068	0.0000	undef	25
Uterus_allgemein	0.0000	0.0000	undef	undef	
Brust-Hyperplasie	0.0000				
Prostata-Hyperplasie	0.0089				
Samenblase	0.0178				
Sinnesorgane	0.0000				30
Weisse_Blutkoerperchen	0.0069				
Zervix	0.0106				

FOETUS				
	%Haeufigkeit			
Entwicklung	0.0000			35
Gastrointestinal	0.0111			
Gehirn	0.0000			
Haematopoetisch	0.0039			
Haut	0.2513			
Hepatisch	0.0000			40
Herz-Blutgefuesse	0.0071			
Lunge	0.0036			
Nebenniere	0.0000			
Niere	0.0062			
Placenta	0.0061			45
Prostata	0.0249			
Sinnesorgane	0.0000			

NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN				
	%Haeufigkeit			
Brust	0.0136			50
Eierstock_n	0.0000			
Eierstock_t	0.0152			
Endokrines Gewebe	0.0000			
Foetal	0.0041			
Gastrointestinal	0.0122			55
Haematopoetisch	0.0000			
Haut-Muskel	0.0130			
Hoden	0.0077			
Lunge	0.0000			
Nerven	0.0040			60
Prostata	0.0137			
Sinnesorgane	0.0000			
Uterus_n	0.0000			65

	NORMAL	TUMOR	Verhaeltnisse	
	%Haeufigkeit	%Haeufigkeit	N/T	T/N
5	Blase 0.0039	0.0204	0.1907	5.2444
	Brust 0.0141	0.0282	0.4991	2.0038
	Duennndarm 0.0061	0.0000	undef	0.0000
	Eierstock 0.0090	0.0312	0.2878	3.4745
	Endokrines_Gewebe 0.0153	0.0176	0.8733	1.1451
10	Gastrointestinal 0.0192	0.0231	0.8283	1.2072
	Gehirn 0.0222	0.0113	1.9635	0.5093
	Haematopoetisch 0.0094	0.0000	undef	0.0000
	Haut 0.0110	0.0000	undef	0.0000
	Hepatisch 0.0048	0.0129	0.3676	2.7200
	Herz 0.0159	0.0000	undef	0.0000
15	Hoden 0.0000	0.0234	0.0000	undef
	Lunge 0.0135	0.0123	1.1007	0.9085
	Magen-Speiseroehre 0.0097	0.0153	0.6303	1.5866
	Muskel-Skelett 0.0188	0.0300	0.6282	1.5918
	Niere 0.0163	0.0205	0.7930	1.2610
20	Pankreas 0.0083	0.0221	0.3739	2.6743
	Penis 0.0090	0.0267	0.3369	2.9678
	Prostata 0.0109	0.0043	2.5591	0.3908
	Uterus_Endometrium 0.0068	0.0000	undef	0.0000
	Uterus_Myometrium 0.0076	0.0204	0.3741	2.6732
25	Uterus_allgemein 0.0102	0.0954	0.1067	9.3678
	Brust-Hyperplasie 0.0032			
	Prostata-Hyperplasie 0.0119			
	Samenblase 0.0089			
	Sinnesorgane 0.0235			
30	Weisse_Blutkoerperchen 0.0087			
	Zervix 0.0000			

	FOETUS
	%Haeufigkeit
35	Entwicklung 0.0000
	Gastrointestinal 0.0111
	Gehirn 0.0000
	Haematopoetisch 0.0039
40	Haut 0.0000
	Hepatisch 0.0260
	Herz-Blutgefuesse 0.0107
	Lunge 0.0072
	Nebenniere 0.0000
45	Niere 0.0124
	Placenta 0.0000
	Prostata 0.0000
	Sinnesorgane 0.0000

	NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN
	%Haeufigkeit
50	Brust 0.0000
	Eierstock_n 0.1595
	Eierstock_t 0.0203
55	Endokrines_Gewebe 0.0000
	Foetal 0.0058
	Gastrointestinal 0.0488
	Haematopoetisch 0.0114
60	Haut-Muskel 0.0032
	Hoden 0.0154
	Lunge 0.0164
	Nerven 0.0060
	Prostata 0.0068
	Sinnesorgane 0.0000
65	Uterus_n 0.0083

## DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 33

	NORMAL %Haeufigkeit	TUMOR %Haeufigkeit	Verhaeltnisse		
			N/T	T/N	
Blase	0.0429	0.2173	0.1974	5.0656	5
Brust	0.0409	0.0357	1.1462	0.8725	
Duenn darm	0.1870	0.0662	2.8269	0.3537	
Eierstock	0.0030	0.0833	0.0360	27.7957	
Endokrines_Gewebe	0.0034	0.0050	0.6792	1.4722	10
Gastrointestinal	0.2740	0.0740	3.7016	0.2702	
Gehirn	0.0007	0.0216	0.0343	29.1683	
Haematopoetisch	0.2165	0.0000	undef	0.0000	
Haut	0.0220	0.0000	undef	0.0000	15
Hepatisch	0.0095	0.3429	0.0277	36.0397	
Herz	0.0074	0.0275	0.2698	3.7059	
Hoden	0.0000	0.0585	0.0000	undef	
Lunge	0.1039	0.1738	0.5977	1.6731	20
Magen-Speiserohre	0.0387	0.0920	0.4202	2.3799	
Muskel-Skelett	0.0994	0.0420	2.3660	0.4227	
Niere	0.0054	0.0753	0.0721	13.8707	
Pankreas	0.0231	0.1160	0.1994	5.0142	25
Penis	0.0000	0.0000	undef	undef	
Prostata	0.0262	0.0000	undef	0.0000	
Uterus_Endometrium	0.0068	0.0000	undef	0.0000	
Uterus_Myometrium	0.0076	0.0000	undef	0.0000	30
Uterus_allgemein	0.0051	0.0000	undef	0.0000	
Brust-Hyperplasie	0.0512				
Prostata-Hyperplasie	0.0178				
Samenblase	0.0089				35
Sinnesorgane	0.0118				
Weisse_Blutkoerperchen	0.0000				
Zervix	0.0319				

FOETUS				
	%Haeufigkeit			
Entwicklung	0.0000			35
Gastrointestinal	0.0000			
Gehirn	0.0000			
Haematopoetisch	0.0079			
Haut	0.0000			40
Hepatisch	0.0000			
Herz-Blutgefuesse	0.0000			
Lunge	0.0000			
Nebenniere	0.0000			45
Niere	0.0000			
Placenta	0.0000			
Prostata	0.0000			
Sinnesorgane	0.0000			50

NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN				
	%Haeufigkeit			
Brust	0.0612			55
Eierstock_n	0.0000			
Eierstock_t	0.0152			
Endokrines_Gewebe	0.0000			
Foetal	0.0012			60
Gastrointestinal	0.0366			
Haematopoetisch	0.0000			
Haut-Muskel	0.0000			
Hoden	0.0000			65
Lunge	0.0000			
Nerven	0.0000			
Prostata	0.0000			
Sinnesorgane	0.0000			65
Uterus_n	0.0000			



# DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 34

		NORMAL	TUMOR	Verhaeltnisse	
		%Haeufigkeit	%Haeufigkeit	N/T	T/N
5	Blase	0.0117	0.0588	0.1990	5.0259
	Brust	0.0102	0.0094	1.0888	0.9184
	Duenn darm	0.0153	0.0331	0.4634	2.1579
	Eierstock	0.0539	0.0130	4.1445	0.2413
	Endokrines Gewebe	0.0187	0.0075	2.4906	0.4015
10	Gastrointestinal	0.0134	0.0093	1.4496	0.6898
	Gehirn	0.0133	0.0123	1.0799	0.9260
	Haematopoetisch	0.0134	0.0000	undef	0.0000
	Haut	0.0147	0.0000	undef	0.0000
	Hepatisch	0.0000	0.0129	0.0000	undef
15	Herz	0.0170	0.0137	1.2336	0.8107
	Hoden	0.0173	0.0234	0.7380	1.3551
	Lunge	0.0062	0.0020	3.0482	0.3281
	Magen-Speiserohre	0.0676	0.0000	undef	0.0000
	Muskel-Skelett	0.0171	0.0180	0.9518	1.0506
20	Niere	0.0109	0.0274	0.3965	2.5219
	Pankreas	0.0066	0.0110	0.5983	1.6714
	Penis	0.0269	0.0533	0.5054	1.9786
	Prostata	0.0327	0.0213	1.5354	0.6513
	Uterus_Endometrium	0.0068	0.0000	undef	0.0000
	Uterus_Myometrium	0.0534	0.0000	undef	0.0000
25	Uterus_allgemein	0.0306	0.0000	undef	0.0000
	Brust-Hyperplasie	0.0000			
	Prostata-Hyperplasie	0.0030			
	Samenblase	0.0000			
	Sinnesorgane	0.0118			
30	Weisse_Blutkoerperchen	0.0061			
	Zervix	0.0213			

		FOETUS
		%Haeufigkeit
35	Entwicklung	0.0139
	Gastrointestinal	0.0555
	Gehirn	0.0000
	Haematopoetisch	0.0079
40	Haut	0.0000
	Hepatisch	0.0260
	Herz-Blutgefuesse	0.0071
	Lunge	0.0253
	Nebenniere	0.0254
45	Niere	0.0000
	Placenta	0.0485
	Prostata	0.0000
	Sinnesorgane	0.0000

		NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN
		%Haeufigkeit
50	Brust	0.0204
	Eierstock_n	0.0000
	Eierstock_t	0.2430
55	Endokrines_Gewebe	0.0245
	Foetal	0.0338
	Gastrointestinal	0.0122
	Haematopoetisch	0.0171
60	Haut-Muskel	0.0680
	Hoden	0.0077
	Lunge	0.0000
	Nerven	0.0151
	Prostata	0.0342
	Sinnesorgane	0.0000
65	Uterus_n	0.1166

## DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 35

	NORMAL	TUMOR	Verhaeltnisse		
	%Haeufigkeit	%Haeufigkeit	N/T	T/N	
Blase	0.0273	0.1227	0.2225	4.4952	5
Brust	0.0243	0.0263	0.9236	1.0828	
Duenndarm	0.1625	0.0000	undef	0.0000	
Eierstock	0.0000	0.0364	0.0000	undef	
Endokrines Gewebe	0.0000	0.0000	undef	undef	
Gastrointestinal	0.2146	0.0093	23.1935	0.0431	10
Gehirn	0.0000	0.0062	0.0000	undef	
Haematopoetisch	0.1203	0.0000	undef	0.0000	
Haut	0.0441	0.0000	undef	0.0000	
Hepatisch	0.0000	0.1229	0.0000	undef	
Herz	0.0021	0.0000	undef	0.0000	15
Hoden	0.0000	0.0935	0.0000	undef	
Lunge	0.0353	0.0900	0.3926	2.5473	
Magen-Speiserohre	0.0000	0.0077	0.0000	undef	
Muskel-Skelett	0.0857	0.0120	7.1388	0.1401	
Niere	0.0081	0.0753	0.1081	9.2471	20
Pankreas	0.0529	0.1381	0.3829	2.6116	
Penis	0.0000	0.0000	undef	undef	
Prostata	0.0109	0.0064	1.7060	0.5862	
Uterus_Endometrium	0.0000	0.0000	undef	undef	
Uterus_Myometrium	0.0000	0.0000	undef	undef	25
Uterus_allgemein	0.0000	0.0000	undef	undef	
Brust-Hyperplasie	0.0160				
Prostata-Hyperplasie	0.0000				
Samenblase	0.0000				
Sinnesorgane	0.0118				30
Weisse_Blutkoerperchen	0.0009				
Zervix	0.0106				

FOETUS				
	%Haeufigkeit			
Entwicklung	0.0000			35
Gastrointestinal	0.0000			
Gehirn	0.0000			
Haematopoetisch	0.0039			
Haut	0.0000			
Hepatisch	0.0000			40
Herz-Blutgefuesse	0.0000			
Lunge	0.0000			
Nebenniere	0.0000			
Niere	0.0000			
Placenta	0.0000			45
Prostata	0.0000			
Sinnesorgane	0.0000			

NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN				
	%Haeufigkeit			
Brust	0.0000			50
Eierstock_n	0.0000			
Eierstock_t	0.0000			
Endokrines_Gewebe	0.0000			
Foetal	0.0000			
Gastrointestinal	0.0000			55
Haematopoetisch	0.0000			
Haut-Muskel	0.0000			
Hoden	0.0000			
Lunge	0.0082			
Nerven	0.0000			60
Prostata	0.0000			
Sinnesorgane	0.0000			
Uterus_n	0.0000			

## Elektronischer Northern für SEQ. ID. NO: 36

40

# DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 37

	NORMAL	TUMOR	Verhaeltnisse		
	%Haeufigkeit	%Haeufigkeit	N/T	T/N	
Blase	0.0078	0.0332	0.2347	4.2611	5
Brust	0.0026	0.0000	undef	0.0000	
Duennndarm	0.0000	0.0000	undef	undef	
Eierstock	0.0000	0.0000	undef	undef	10
Endokrines Gewebe	0.0000	0.0000	undef	undef	
Gastrointestinal	0.0115	0.0000	undef	0.0000	
Gehirn	0.0000	0.0000	undef	undef	
Haematopoetisch	0.0147	0.0000	undef	0.0000	
Haut	0.0000	0.0000	undef	undef	15
Hepatisch	0.0000	0.0000	undef	undef	
Herz	0.0000	0.0000	undef	undef	
Hoden	0.0000	0.0000	undef	undef	
Lunge	0.0021	0.0020	1.0161	0.9842	
Magen-Speiserohre	0.0000	0.0000	undef	undef	20
Muskel-Skelett	0.0069	0.0000	undef	0.0000	
Niere	0.0000	0.0000	undef	undef	
Pankreas	0.0017	0.0607	0.0272	36.7712	
Penis	0.0000	0.0000	undef	undef	
Prostata	0.0000	0.0000	undef	undef	25
Uterus_Endometrium	0.0000	0.0000	undef	undef	
Uterus_Myometrium	0.0000	0.0000	undef	undef	
Uterus_allgemein	0.0000	0.0000	undef	undef	
Brust-Hyperplasie	0.0000				
Prostata-Hyperplasie	0.0000				30
Samenblase	0.0000				
Sinnesorgane	0.0000				
Weisse_Blutkoerperchen	0.0009				
Zervix	0.0000				

	FOETUS	
	%Haeufigkeit	
Entwicklung	0.0000	35
Gastrointestinal	0.0000	
Gehirn	0.0000	
Haematopoetisch	0.0039	40
Haut	0.0000	
Hepatisch	0.0000	
Herz-Blutgefuesse	0.0000	
Lunge	0.0000	
Nebenniere	0.0000	45
Niere	0.0000	
Placenta	0.0000	
Prostata	0.0000	
Sinnesorgane	0.0000	

	NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN	
	%Haeufigkeit	
Brust	0.0000	50
Eierstock_n	0.0000	
Eierstock_t	0.0000	
Endokrines Gewebe	0.0000	55
Foetal	0.0000	
Gastrointestinal	0.0000	
Haematopoetisch	0.0000	60
Haut-Muskel	0.0000	
Hoden	0.0000	
Lunge	0.0000	
Nerven	0.0000	
Prostata	0.0000	65
Sinnesorgane	0.0000	
Uterus_n	0.0000	

## DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 38

		NORMAL	TUMOR	Verhaeltnisse	
		%Haeufigkeit	%Haeufigkeit	N/T	T/N
5	Blase	0.0624	0.2607	0.2393	4.1791
	Brust	0.0051	0.0282	0.1815	5.5104
	Duennndarm	0.0337	0.0000	undef	0.0000
	Eierstock	0.0030	0.0937	0.0320	31.2702
	Endokrines_Gewebe	0.0034	0.0025	1.3585	0.7361
10	Gastrointestinal	0.1360	0.0694	1.9604	0.5101
	Gehirn	0.0015	0.0318	0.0464	21.5290
	Haematopoetisch	0.1337	0.0000	undef	0.0000
	Haut	0.0147	0.0000	undef	0.0000
	Hepatisch	0.0000	0.4594	0.0000	undef
15	Herz	0.0127	0.0137	0.9252	1.0809
	Hoden	0.0000	0.1754	0.0000	undef
	Lunge	0.0540	0.2086	0.2590	3.8610
	Magen-Speiserohre	0.0193	0.0767	0.2521	3.9666
	Muskel-Skelett	0.1747	0.0240	7.2815	0.1373
20	Niere	0.0190	0.0205	0.9252	1.0808
	Pankreas	0.0611	0.0773	0.7906	1.2649
	Penis	0.0030	0.0000	undef	0.0000
	Prostata	0.0196	0.0021	9.2126	0.1085
	Uterus_Endometrium	0.0135	0.0000	undef	0.0000
	Uterus_Myometrium	0.0000	0.0000	undef	undef
25	Uterus_allgemein	0.0000	0.0000	undef	undef
	Brust-Hyperplasie	0.0064			
	Prostata-Hyperplasie	0.0089			
	Samenblase	0.0089			
	Sinnesorgane	0.0353			
30	Weisse_Blutkoerperchen	0.0009			
	Zervix	0.0426			

		FOETUS
		%Haeufigkeit
35	Entwicklung	0.0000
	Gastrointestinal	0.0000
	Gehirn	0.0000
	Haematopoetisch	0.0000
40	Haut	0.0000
	Hepatisch	0.0000
	Herz-Blutgefuesse	0.0000
	Lunge	0.0000
	Nebenniere	0.0000
45	Niere	0.0000
	Placenta	0.0061
	Prostata	0.0000
	Sinnesorgane	0.0000

		NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN
		%Haeufigkeit
50	Brust	0.0000
	Eierstock_n	0.0000
	Eierstock_t	0.0101
55	Endokrines_Gewebe	0.0000
	Foetal	0.0000
	Gastrointestinal	0.0000
	Haematopoetisch	0.0000
60	Haut-Muskel	0.0000
	Hoden	0.0000
	Lunge	0.0000
	Nerven	0.0010
	Prostata	0.0000
	Sinnesorgane	0.0000
65	Uterus_n	0.0000

## DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 39

	NORMAL	TUMOR	Verhaeltnisse		
	%Haeufigkeit	%Haeufigkeit	N/T	T/N	
Blase	0.0702	0.2761	0.2542	3.9333	5
Brust	0.0614	0.0620	0.9899	1.0102	
Duenn darm	0.2024	0.1489	1.3594	0.7356	
Eierstock	0.0060	0.0442	0.1354	7.3832	
Endokrines_Gewebe	0.0255	0.0025	10.1887	0.0981	
Gastrointestinal	0.4138	0.0740	5.5913	0.1788	10
Gehirn	0.0044	0.0431	0.1029	9.7228	
Haematopoetisch	0.5774	0.0000	undef	0.0000	
Haut	0.0257	0.0000	undef	0.0000	
Hepatisch	0.0285	0.3364	0.0848	11.7866	
Herz	0.0095	0.0412	0.2313	4.3235	15
Hoden	0.0000	0.2689	0.0000	undef	
Lunge	0.1735	0.2678	0.6476	1.5441	
Magen-Speiserohre	0.0870	0.1917	0.4538	2.2037	
Muskel-Skelett	0.2193	0.0180	12.1835	0.0821	
Niere	0.0136	0.1164	0.1166	8.5746	20
Pankreas	0.0677	0.1822	0.3717	2.6906	
Penis	0.0030	0.1333	0.0225	44.5175	
Prostata	0.0327	0.0085	3.8386	0.2605	
Uterus_Endometrium	0.0000	0.0000	undef	undef	
Uterus_Myometrium	0.0000	0.0000	undef	undef	25
Uterus_allgemein	0.0051	0.0000	undef	0.0000	
Brust-Hyperplasie	0.0927				
Prostata-Hyperplasie	0.0178				
Samenblase	0.0267				
Sinnesorgane	0.2823				30
Weisse_Blutkoerperchen	0.0087				
Zervix	0.0426				

FOETUS					
%Haeufigkeit					
Entwicklung	0.0000				35
Gastrointestinal	0.0056				
Gehirn	0.0000				
Haematopoetisch	0.0157				
Haut	0.0000				
Hepatisch	0.0000				40
Herz-Blutgefuesse	0.0000				
Lunge	0.0000				
Nebenniere	0.0254				
Niere	0.0000				
Placenta	0.0000				45
Prostata	0.0000				
Sinnesorgane	0.0000				

NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN					
%Haeufigkeit					
Brust	0.4082				50
Eierstock_n	0.0000				
Eierstock_t	0.3493				
Endokrines_Gewebe	0.0000				
Foetal	0.0082				
Gastrointestinal	0.1953				55
Haematopoetisch	0.0000				
Haut-Muskel	0.0000				
Hoden	0.0000				
Lunge	0.0491				
Nerven	0.0030				60
Prostata	0.0000				
Sinnesorgane	0.0000				
Uterus_n	0.0000				

## DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 40

		NORMAL	TUMOR	Verhaeltnisse	
		%Haeufigkeit	%Haeufigkeit	N/T	T/N
5	Blase	0.0195	0.0690	0.2825	3.5400
	Brust	0.0166	0.0320	0.5204	1.9216
	Duennndarm	0.0031	0.0000	undef	0.0000
	Eierstock	0.0150	0.0130	1.1513	0.8686
	Endokrines_Gewebe	0.0085	0.0100	0.8491	1.1778
10	Gastrointestinal	0.0019	0.0093	0.2071	4.8289
	Gehirn	0.0067	0.0442	0.1507	6.6362
	Haematopoetisch	0.0187	0.0000	undef	0.0000
	Haut	0.0073	0.0000	undef	0.0000
	Hepatisch	0.0238	0.0194	1.2255	0.8160
	Herz	0.0625	0.1512	0.4135	2.4182
15	Hoden	0.0345	0.0117	2.9518	0.3388
	Lunge	0.0322	0.0286	1.1249	0.8889
	Magen-Speiserohre	0.0000	0.0307	0.0000	undef
	Muskel-Skelett	0.0668	0.1260	0.5303	1.8857
	Niere	0.0190	0.0342	0.5551	1.8014
20	Pankreas	0.0050	0.1160	0.0427	23.3998
	Penis	0.0299	0.0000	undef	0.0000
	Prostata	0.0131	0.0170	0.7677	1.3026
	Uterus_Endometrium	0.0068	0.0528	0.1280	7.8106
	Uterus_Myometrium	0.0305	0.0204	1.4964	0.6683
25	Uterus_allgemein	0.0357	0.0000	undef	0.0000
	Brust-Hyperplasie	0.0000			
	Prostata-Hyperplasie	0.0208			
	Samenblase	0.0356			
	Sinnesorgane	0.0588			
30	Weisse_Blutkoerperchen	0.0000			
	Zervix	0.0319			

		FOETUS
		%Haeufigkeit
35	Entwicklung	0.0000
	Gastrointestinal	0.0167
	Gehirn	0.0000
	Haematopoetisch	0.0118
40	Haut	0.0000
	Hepatisch	0.0520
	Herz-Blutgefuesse	0.0107
	Lunge	0.0325
	Nebenniere	0.0000
45	Niere	0.0494
	Placenta	0.0909
	Prostata	0.0000
	Sinnesorgane	0.0000

		NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN
		%Haeufigkeit
50	Brust	0.0340
	Eierstock_n	0.0000
	Eierstock_t	0.0203
55	Endokrines_Gewebe	0.0490
	Foetal	0.0297
	Gastrointestinal	0.0000
	Haematopoetisch	0.0000
	Haut-Muskel	0.0000
60	Hoden	0.0154
	Lunge	0.0082
	Nerven	0.0030
	Prostata	0.0000
	Sinnesorgane	0.0000
65	Uterus_n	0.0000

## DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 41

	NORMAL	TUMOR	Verhaeltnisse	
	%Haeufigkeit	%Haeufigkeit	N/T T/N	
Blase	0.0390	0.1355	0.2878 3.4744	5
Brust	0.0269	0.0564	0.4764 2.0992	
Duennndarm	0.1932	0.0000	undef 0.0000	
Eierstock	0.0000	0.0312	0.0000 undef	
Endokrines_Gewebe	0.0017	0.0025	0.6792 1.4722	
Gastrointestinal	0.2203	0.0509	4.3300 0.2309	10
Gehirn	0.0000	0.0103	0.0000 undef	
Haematopoetisch	0.0882	0.0000	undef 0.0000	
Haut	0.0184	0.0000	undef 0.0000	
Hepatisch	0.0000	0.0518	0.0000 undef	
Herz	0.0095	0.0137	0.6939 1.4412	15
Hoden	0.0000	0.0702	0.0000 undef	
Lunge	0.0478	0.1472	0.3246 3.0809	
Magen-Speiserohre	0.0193	0.0153	1.2605 0.7933	
Muskel-Skelett	0.1011	0.0120	8.4237 0.1187	
Niere	0.0054	0.0548	0.0991 10.0878	20
Pankreas	0.0099	0.0552	0.1795 5.5714	
Penis	0.0000	0.2933	0.0000 undef	
Prostata	0.0153	0.0021	7.1654 0.1396	
Uterus_Endometrium	0.0068	0.0000	undef 0.0000	
Uterus_Myometrium	0.0000	0.0068	0.0000 undef	25
Uterus_allgemein	0.0000	0.0000	undef undef	
Brust-Hyperplasie	0.0192			
Prostata-Hyperplasie	0.0119			
Samenblase	0.0000			
Sinnesorgane	0.0000			
Weisse_Blutkoerperchen	0.0009			30
Zervix	0.0213			

	FOETUS	
	%Haeufigkeit	
Entwicklung	0.0000	35
Gastrointestinal	0.0000	
Gehirn	0.0000	
Haematopoetisch	0.0079	
Haut	0.0000	40
Hepatisch	0.0000	
Herz-Blutgefuesse	0.0000	
Lunge	0.0000	
Nebenniere	0.0000	
Niere	0.0000	45
Placenta	0.0000	
Prostata	0.0000	
Sinnesorgane	0.0000	

	NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN	
	%Haeufigkeit	
Brust	0.0000	50
Eierstock_n	0.0000	
Eierstock_t	0.0000	
Endokrines_Gewebe	0.0000	55
Foetal	0.0000	
Gastrointestinal	0.0000	
Haematopoetisch	0.0000	
Haut-Muskel	0.0000	60
Hoden	0.0000	
Lunge	0.0000	
Nerven	0.0000	
Prostata	0.0000	
Sinnesorgane	0.0000	
Uterus_n	0.0000	65



## DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 42

	NORMAL	TUMOR	Verhaeltnisse	
	%Haeufigkeit	%Haeufigkeit	N/T	T/N
5	Blase 0.0273	0.0895	0.3051	3.2777
	Brust 0.0665	0.0808	0.8230	1.2151
	Duennndarm 0.0429	0.0000	undef	0.0000
	Eierstock 0.0539	0.0234	2.3025	0.4343
	Endokrines_Gewebe 0.0630	0.0978	0.6444	1.5518
10	Gastrointestinal 0.0441	0.0324	1.3608	0.7348
	Gehirn 0.0554	0.1006	0.5510	1.8149
	Haematopoetisch 0.0454	0.0379	1.1998	0.8335
	Haut 0.0257	0.0000	undef	0.0000
	Hepatisch 0.0381	0.0453	0.8403	1.1900
	Herz 0.0435	0.0825	0.5268	1.8981
15	Hoden 0.0575	0.0000	undef	0.0000
	Lunge 0.1008	0.0552	1.8252	0.5479
	Magen-Speiserohre 0.0580	0.0997	0.5818	1.7188
	Muskel-Skelett 0.0976	0.0660	1.4797	0.6758
	Niere 0.0516	0.0890	0.5795	1.7255
20	Pankreas 0.0248	0.0773	0.3205	3.1200
	Penis 0.0599	0.1066	0.5616	1.7807
	Prostata 0.0567	0.0766	0.7393	1.3527
	Uterus_Endometrium 0.0405	0.1055	0.3841	2.6035
	Uterus_Myometrium 0.0534	0.0475	1.1223	0.8911
25	Uterus_allgemein 0.0866	0.0000	undef	0.0000
	Brust-Hyperplasie 0.0416			
	Prostata-Hyperplasie 0.0654			
	Samenblase 0.0712			
	Sinnesorgane 0.0823			
30	Weisse_Blutkoerperchen 0.1110			
	Zervix 0.0319			

	FOETUS
	%Haeufigkeit
35	Entwicklung 0.0139
	Gastrointestinal 0.0361
	Gehirn 0.0125
	Haematopoetisch 0.0433
40	Haut 0.0000
	Hepatisch 0.0000
	Herz-Blutgefuesse 0.0071
	Lunge 0.0253
	Nebenniere 0.0254
45	Niere 0.0432
	Placenta 0.0364
	Prostata 0.0499
	Sinnesorgane 0.0126

	NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN
	%Haeufigkeit
50	Brust 0.1020
	Eierstock_n 0.1595
	Eierstock_t 0.0709
55	Endokrines_Gewebe 0.0000
	Foetal 0.0635
	Gastrointestinal 0.0000
	Haematopoetisch 0.0114
60	Haut-Muskel 0.0680
	Hoden 0.0463
	Lunge 0.0328
	Nerven 0.0351
	Prostata 0.0342
	Sinnesorgane 0.0464
65	Uterus_n 0.0083

## DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 43

	NORMAL	TUMOR	Verhaeltnisse	
	%Haeufigkeit	%Haeufigkeit	N/T T/N	
Blase	0.0195	0.0639	0.3051 3.2777	5
Brust	0.0345	0.0470	0.7350 1.3606	
Duennndarm	0.0399	0.1985	0.2008 4.9797	
Eierstock	0.0150	0.0676	0.2214 4.5168	
Endokrines_Gewebe	0.0238	0.0878	0.2717 3.6805	
Gastrointestinal	0.0862	0.1064	0.8103 1.2341	10
Gehirn	0.0067	0.0277	0.2400 4.1669	
Haematopoetisch	0.0080	0.0000	undef 0.0000	
Haut	0.0587	0.0000	undef 0.0000	
Hepatisch	0.0190	0.0323	0.5882 1.7000	
Herz	0.0011	0.0962	0.0110 90.7941	15
Hoden	0.0000	0.0234	0.0000 undef	
Lunge	0.0062	0.0164	0.3810 2.6245	
Magen-Speiserohre	0.0387	0.3450	0.1120 8.9248	
Muskel-Skelett	0.0000	0.0360	0.0000 undef	
Niere	0.0760	0.1643	0.4626 2.1617	20
Pankreas	0.0677	0.0276	2.4530 0.4077	
Penis	0.0090	0.0533	0.1685 5.9357	
Prostata	0.0109	0.0255	0.4265 2.3446	
Uterus_Endometrium	0.0270	0.0000	undef 0.0000	
Uterus_Myometrium	0.0076	0.0272	0.2806 3.5642	25
Uterus_allgemein	0.0000	0.4771	0.0000 undef	
Brust-Hyperplasie	0.0576			
Prostata-Hyperplasie	0.0119			
Samenblase	0.1068			
Sinnesorgane	0.0235			
Weisse_Blutkoerperchen	0.0061			30
Zervix	0.0319			

	FOETUS	
	%Haeufigkeit	
Entwicklung	0.0000	35
Gastrointestinal	0.0111	
Gehirn	0.0813	
Haematopoetisch	0.0000	
Haut	0.0000	40
Hepatisch	0.0000	
Herz-Blutgefuesse	0.0000	
Lunge	0.0145	
Nebenniere	0.0000	
Niere	0.0309	45
Placenta	0.0121	
Prostata	0.0249	
Sinnesorgane	0.0000	

	NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN	
	%Haeufigkeit	
Brust	0.0136	
Eierstock_n	0.0000	
Eierstock_t	0.0101	
Endokrines_Gewebe	0.0000	55
Foetal	0.0122	
Gastrointestinal	0.4149	
Haematopoetisch	0.0000	
Haut-Muskel	0.0000	60
Hoden	0.0154	
Lunge	0.0573	
Nerven	0.0040	
Prostata	0.0068	
Sinnesorgane	0.0000	
Uterus_n	0.0000	65

		NORMAL	TUMOR	Verhaeltnisse	
		%Haeufigkeit	%Haeufigkeit	N/T	T/N
5	Blase	0.0156	0.0511	0.3051	3.2777
	Brust	0.0166	0.0320	0.5204	1.9216
	Duennndarm	0.0828	0.0000	undef	0.0000
	Eierstock	0.0030	0.0416	0.0720	13.8979
	Endokrines_Gewebe	0.0017	0.0050	0.3396	2.9444
10	Gastrointestinal	0.1226	0.0278	4.4178	0.2264
	Gehirn	0.0000	0.0123	0.0000	undef
	Haematopoetisch	0.0481	0.0000	undef	0.0000
	Haut	0.0294	0.0000	undef	0.0000
	Hepatisch	0.0000	0.1294	0.0000	undef
15	Herz	0.0032	0.0137	0.2313	4.3235
	Hoden	0.0000	0.1403	0.0000	undef
	Lunge	0.0447	0.1390	0.3213	3.1128
	Magen-Speiserohre	0.0097	0.0000	undef	0.0000
	Muskel-Skelett	0.0497	0.0060	8.2810	0.1208
20	Niere	0.0054	0.0685	0.0793	12.6097
	Pankreas	0.0215	0.0166	1.2963	0.7714
	Penis	0.0030	0.0000	undef	0.0000
	Prostata	0.0044	0.0043	1.0236	0.9769
	Uterus_Endometrium	0.0000	0.0000	undef	undef
25	Uterus_Myometrium	0.0000	0.0000	undef	undef
	Uterus_allgemein	0.0000	0.0000	undef	undef
	Brust-Hyperplasie	0.0352			
	Prostata-Hyperplasie	0.0030			
	Samenblase	0.0000			
	Sinnesorgane	0.0000			
30	Weisse_Blutkoerperchen	0.0000			
	Zervix	0.0106			

		FOETUS
		%Haeufigkeit
35	Entwicklung	0.0000
	Gastrointestinal	0.0000
	Gehirn	0.0000
	Haematopoetisch	0.0079
40	Haut	0.0000
	Hepatisch	0.0000
	Herz-Blutgefuesse	0.0000
	Lunge	0.0000
	Nebenniere	0.0000
45	Niere	0.0000
	Placenta	0.0000
	Prostata	0.0000
	Sinnesorgane	0.0000

		NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN
		%Haeufigkeit
50	Brust	0.0136
	Eierstock_n	0.0000
	Eierstock_t	0.0000
55	Endokrines_Gewebe	0.0000
	Foetal	0.0000
	Gastrointestinal	0.0000
	Haematopoetisch	0.0000
60	Haut-Muskel	0.0000
	Hoden	0.0000
	Lunge	0.0000
	Nerven	0.0000
	Prostata	0.0000
65	Sinnesorgane	0.0000
	Uterus_n	0.0000

## DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 45

	NORMAL	TUMOR	Verhaeltnisse		
	%Haeufigkeit	%Haeufigkeit	N/T	T/N	
Blase	0.0117	0.0383	0.3051	3.2777	5
Brust	0.0179	0.0207	0.8661	1.1546	
Duennndarm	0.0245	0.0165	1.4830	0.6743	
Eierstock	0.0329	0.0156	2.1106	0.4738	
Endokrines_Gewebe	0.0221	0.0326	0.6792	1.4722	
Gastrointestinal	0.0153	0.0139	1.1045	0.9054	10
Gehirn	0.0296	0.0288	1.0285	0.9723	
Haematopoetisch	0.0187	0.0379	0.4940	2.0241	
Haut	0.0257	0.0000	undef	0.0000	
Hepatisch	0.0048	0.0129	0.3676	2.7200	
Herz	0.0201	0.0550	0.3662	2.7306	15
Hoden	0.0058	0.0117	0.4920	2.0326	
Lunge	0.0218	0.0225	0.9699	1.0311	
Magen-Speiseroehre	0.0387	0.0383	1.0084	0.9916	
Muskel-Skelett	0.0171	0.0240	0.7139	1.4008	
Niere	0.0190	0.0068	2.7756	0.3603	20
Pankreas	0.0116	0.0221	0.5235	1.9102	
Penis	0.0090	0.0000	undef	0.0000	
Prostata	0.0327	0.0319	1.0236	0.9769	
Uterus_Endometrium	0.0068	0.0000	undef	0.0000	
Uterus_Myometrium	0.0000	0.0136	0.0000	undef	25
Uterus_allgemein	0.0560	0.0000	undef	0.0000	
Brust-Hyperplasie	0.0032				
Prostata-Hyperplasie	0.0208				
Samenblase	0.0178				
Sinnesorgane	0.0118				
Weisse_Blutkoerperchen	0.0260				30
Zervix	0.0000				

	FOETUS	
	%Haeufigkeit	
Entwicklung	0.0000	35
Gastrointestinal	0.0139	
Gehirn	0.0313	
Haematopoetisch	0.0039	
Haut	0.0000	40
Hepatisch	0.0000	
Herz-Blutgefuesse	0.0071	
Lunge	0.0217	
Nebenniere	0.0000	
Niere	0.0000	45
Placenta	0.0424	
Prostata	0.0249	
Sinnesorgane	0.0000	

	NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN	
	%Haeufigkeit	
Brust	0.0408	50
Eierstock_n	0.0000	
Eierstock_t	0.0405	
Endokrines_Gewebe	0.0245	55
Foetal	0.0087	
Gastrointestinal	0.0000	
Haematopoetisch	0.0057	
Haut-Muskel	0.0097	60
Hoden	0.0231	
Lunge	0.0082	
Nerven	0.0251	
Prostata	0.0205	
Sinnesorgane	0.0000	
Uterus_n	0.0125	65

## DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 46

	NORMAL	TUMOR	Verhaeltnisse	
	%Haeufigkeit	%Haeufigkeit	N/T	T/N
5	Blase 0.1170	0.3067	0.3814	2.6222
	Brust 0.3019	0.2387	1.2646	0.7908
	Duennndarm 1.1559	0.0000	undef	0.0000
	Eierstock 0.0120	0.0676	0.1771	5.6460
	Endokrines_Gewebe 0.0034	0.0125	0.2717	3.6805
10	Gastrointestinal 1.2798	0.1804	7.0940	0.1410
	Gehirn 0.0007	0.0380	0.0195	51.3918
	Haematopoetisch 0.4785	0.0000	undef	0.0000
	Haut 0.1322	0.0000	undef	0.0000
	Hepatisch 0.0285	0.4594	0.0621	16.0932
15	Herz 0.0138	0.0275	0.5011	1.9955
	Hoden 0.0000	0.2456	0.0000	undef
	Lunge 0.3625	0.3435	1.0554	0.9475
	Magen-Speiserohre 0.0483	0.1533	0.3151	3.1733
	Muskel-Skelett 0.2124	0.0480	4.4260	0.2259
20	Niere 0.0163	0.1917	0.0850	11.7691
	Pankreas 0.1074	0.4528	0.2371	4.2171
	Penis 0.0000	0.0533	0.0000	undef
	Prostata 0.0806	0.0255	3.1562	0.3168
	Uterus_Endometrium 0.0068	0.0000	undef	0.0000
	Uterus_Myometrium 0.0152	0.0000	undef	0.0000
25	Uterus_allgemein 0.0102	0.0000	undef	0.0000
	Brust-Hyperplasie 0.3805			
	Prostata-Hyperplasie 0.0505			
	Samenblase 0.0356			
	Sinnesorgane 0.5175			
30	Weisse_Blutkoerperchen 0.0078			
	Zervix 0.0213			

	FOETUS
	%Haeufigkeit.
35	Entwicklung 0.0000
	Gastrointestinal 0.0000
	Gehirn 0.0000
	Haematopoetisch 0.0039
40	Haut 0.0000
	Hepatisch 0.0000
	Herz-Blutgefuesse 0.0000
	Lunge 0.0000
	Nebenniere 0.0000
45	Niere 0.0000
	Placenta 0.0303
	Prostata 0.0000
	Sinnesorgane 0.0000

	NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN
	%Haeufigkeit
50	Brust 0.7687
	Eierstock_n 0.0000
	Eierstock_t 0.1924
55	Endokrines_Gewebe 0.0000
	Foetal 0.0029
	Gastrointestinal 0.0976
	Haematopoetisch 0.0000
60	Haut-Muskel 0.0000
	Hoden 0.0000
	Lunge 0.0328
	Nerven 0.0020
	Prostata 0.0000
	Sinnesorgane 0.0000
65	Uterus_n 0.0000

## DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 47

	NORMAL	TUMOR	Verhaeltnisse		
	%Haeufigkeit	%Haeufigkeit	N/T	T/N	
Blase	0.0195	0.0486	0.4014	2.4911	5
Brust	0.0550	0.1297	0.4241	2.3580	
Duennndarm	0.0123	0.1654	0.0741	13.4866	
Eierstock	0.0449	0.0728	0.6167	1.6214	
Endokrines_Gewebe	0.0119	0.0075	1.5849	0.6309	
Gastrointestinal	0.0153	0.1018	0.1506	6.6398	10
Gehirn	0.0037	0.0359	0.1029	9.7228	
Haematopoetisch	0.0134	0.0000	undef	0.0000	
Haut	0.0404	0.0847	0.4765	2.0985	
Hepatisch	0.0285	0.0323	0.8824	1.1333	
Herz	0.0233	0.0825	0.2827	3.5374	15
Hoden	0.0460	0.0351	1.3119	0.7622	
Lunge	0.0551	0.0491	1.1219	0.8913	
Magen-Speiserohre	0.0000	0.0613	0.0000	undef	
Muskel-Skelett	0.1148	0.3120	0.3679	2.7180	
Niere	0.0271	0.0137	1.9826	0.5044	20
Pankreas	0.0116	0.0884	0.1309	7.6408	
Penis	0.0180	0.0533	0.3369	2.9678	
Prostata	0.0087	0.0319	0.2730	3.6634	
Uterus_Endometrium	0.0270	0.0000	undef	0.0000	25
Uterus_Myometrium	0.0610	0.1902	0.3206	3.1187	
Uterus_allgemein	0.0509	0.0000	undef	0.0000	
Brust-Hyperplasie	0.0288				
Prostata-Hyperplasie	0.0089				
Samenblase	0.0178				
Sinnesorgane	0.0118				30
Weisse_Blutkoerperchen	0.0000				
Zervix	0.0532				

	FOETUS	
	%Haeufigkeit	
Entwicklung	0.1809	35
Gastrointestinal	0.1388	
Gehirn	0.0063	
Haematopoetisch	0.0551	
Haut	0.2513	40
Hepatisch	0.0000	
Herz-Blutgefuesse	0.0605	
Lunge	0.1409	
Nebenniere	0.2535	
Niere	0.1235	45
Placenta	0.0364	
Prostata	0.0748	
Sinnesorgane	1.3934	

	NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN	
	%Haeufigkeit	
Brust	0.0204	50
Eierstock_n	0.0000	
Eierstock_t	0.0304	55
Endokrines_Gewebe	0.0000	
Foetal	0.0315	
Gastrointestinal	0.0122	
Haematopoetisch	0.0000	
Haut-Muskel	0.0518	60
Hoden	0.0231	
Lunge	0.0000	
Nerven	0.0040	
Prostata	0.0137	
Sinnesorgane	0.0000	65
Uterus_n	0.0083	

## DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 48

		NORMAL	TUMOR	Verhaeltnisse	
		%Haeufigkeit	%Haeufigkeit	N/T	T/N
5	Blase	0.0624	0.1380	0.4520	2.2125
	Brust	0.0345	0.0564	0.6125	1.6327
	Duenndarm	0.0491	0.0165	2.9659	0.3372
	Eierstock	0.0689	0.0520	1.3239	0.7553
	Endokrines_Gewebe	0.0392	0.0276	1.4202	0.7041
10	Gastrointestinal	0.0460	0.0648	0.7100	1.4084
	Gehirn	0.0333	0.0678	0.4909	2.0372
	Haematopoetisch	0.0374	0.0758	0.4940	2.0241
	Haut	0.0257	0.1695	0.1516	6.5954
	Hepatisch	0.1142	0.0518	2.2059	0.4533
15	Herz	0.0774	0.7010	0.1104	9.0616
	Hoden	0.2589	0.1520	1.7030	0.5872
	Lunge	0.0540	0.0491	1.1007	0.9085
	Magen-Speiserohre	0.1256	0.2070	0.6069	1.6477
	Muskel-Skelett	0.1542	0.2100	0.7343	1.3619
20	Niere	0.0109	0.0959	0.1133	8.8268
	Pankreas	0.0198	0.1270	0.1561	6.4071
	Penis	0.0359	0.0800	0.4493	2.2259
	Prostata	0.0785	0.0511	1.5354	0.6513
	Uterus_Endometrium	0.0338	0.0528	0.6402	1.5621
25	Uterus_Myometrium	0.0457	0.0679	0.6734	1.4851
	Uterus_allgemein	0.0764	0.0954	0.8006	1.2490
	Brust-Hyperplasie	0.0224			
	Prostata-Hyperplasie	0.0624			
	Samenblase	0.0445			
	Sinnesorgane	0.0118			
30	Weisse_Blutkoerperchen	0.0529			
	Zervix	0.0532			

		FOETUS
		%Haeufigkeit.
35	Entwicklung	0.0139
	Gastrointestinal	0.0167
	Gehirn	0.0000
	Haematopoetisch	0.0039
40	Haut	0.0000
	Hepatisch	0.0000
	Herz-Blutgefuesse	0.0249
	Lunge	0.0108
	Nebenniere	0.0000
45	Niere	0.0000
	Placenta	0.3333
	Prostata	0.1995
	Sinnesorgane	0.0000

		NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN
		%Haeufigkeit
50	Brust	0.1156
	Eierstock_n	0.0000
55	Eierstock_t	0.1873
	Endokrines_Gewebe	0.0000
	Foetal	0.0181
	Gastrointestinal	0.0000
	Haematopoetisch	0.0057
60	Haut-Muskel	0.0486
	Hoden	0.0000
	Lunge	0.0328
	Nerven	0.0020
	Prostata	0.0274
65	Sinnesorgane	0.0155
	Uterus_n	0.0541

# DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 49

	NORMAL	TUMOR	Verhaeltnisse	
	%Haeufigkeit	%Haeufigkeit	N/T T/N	
Blase	0.0585	0.1278	0.4576 2.1852	5
Brust	0.0742	0.0507	1.4619 0.6841	
Duenn darm	0.1134	0.0000	undef 0.0000	
Eierstock	0.0659	0.0390	1.6885 0.5922	
Endokrines_Gewebe	0.0324	0.0176	1.8437 0.5424	
Gastrointestinal	0.0690	0.0694	0.9940 1.0060	10
Gehirn	0.0325	0.0893	0.3641 2.7464	
Haematopoetisch	0.0842	0.0000	undef 0.0000	
Haut	0.0404	0.0000	undef 0.0000	
Hepatisch	0.1855	0.0647	2.8677 0.3487	
Herz	0.0435	0.0962	0.4516 2.2145	15
Hoden	0.0230	0.0585	0.3936 2.5408	
Lunge	0.2275	0.1063	2.1396 0.4674	
Magen-Speiserohre	0.0483	0.1993	0.2424 4.1252	
Muskel-Skelett	0.0857	0.0720	1.1898 0.8405	
Niere	0.0706	0.0274	2.5774 0.3880	20
Pankreas	0.0347	0.2927	0.1185 8.4367	
Penis	0.0090	0.0267	0.3369 2.9678	
Prostata	0.0458	0.0319	1.4331 0.6978	
Uterus_Endometrium	0.0270	0.0000	undef 0.0000	
Uterus_Myometrium	0.0610	0.0204	2.9927 0.3341	25
Uterus_allgemein	0.0917	0.0000	undef 0.0000	
Brust-Hyperplasie	0.0416			
Prostata-Hyperplasie	0.0386			
Samenblase	0.0267			
Sinnesorgane	0.1647			
Weisse_Blutkoerperchen	0.2497			30
Zervix	0.0000			
	FOETUS			
	%Haeufigkeit			35
Entwicklung	0.0000			
Gastrointestinal	0.0916			
Gehirn	0.0500			
Haematopoetisch	0.0865			
Haut	0.0000			40
Hepatisch	0.0520			
Herz-Blutgefuesse	0.0178			
Lunge	0.0434			
Nebenniere	0.0000			
Niere	0.0432			45
Placenta	0.1636			
Prostata	0.0000			
Sinnesorgane	0.0000			
	NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN			50
	%Haeufigkeit			
Brust	0.1156			
Eierstock_n	0.0000			
Eierstock_t	0.0354			
Endokrines_Gewebe	0.0490			55
Foetal	0.0571			
Gastrointestinal	0.0854			
Haematopoetisch	0.0057			
Haut-Muskel	0.0097			
Hoden	0.0154			60
Lunge	0.0491			
Nerven	0.0562			
Prostata	0.0615			
Sinnesorgane	0.0000			
Uterus_n	0.1957			65



## DE 198 18 619 A 1

Elektronischer Northern für SEQ. ID. NO: 50

	NORMAL	TUMOR	Verhaeltnisse	
	%Haeufigkeit	%Haeufigkeit	N/T	T/N
5	Blase 0.0546	0.1099	0.4967	2.0135
	Brust 0.0563	0.0489	1.1517	0.8683
	Duennndarm 0.1380	0.0331	4.1708	0.2398
	Eierstock 0.0599	0.0650	0.9210	1.0858
	Endokrines_Gewebe 0.0238	0.0351	0.6792	1.4722
10	Gastrointestinal 0.1322	0.1804	0.7328	1.3647
	Gehirn 0.0229	0.0452	0.5072	1.9714
	Haematopoetisch 0.0241	0.0000	undef	0.0000
	Haut 0.1689	0.0000	undef	0.0000
	Hepatisch 0.0285	0.0518	0.5515	1.8133
15	Herz 0.1166	0.1649	0.7067	1.4150
	Hoden 0.0115	0.0117	0.9839	1.0163
	Lunge 0.1070	0.1329	0.8050	1.2422
	Magen-Speiserohre 0.1450	0.0613	2.3635	0.4231
	Muskel-Skelett 0.0685	0.0240	2.8555	0.3502
20	Niere 0.0570	0.0753	0.7570	1.3210
	Pankreas 0.0165	0.1491	0.1108	9.0256
	Penis 0.0779	0.0267	2.9202	0.3424
	Prostata 0.0610	0.0255	2.3885	0.4187
	Uterus_Endometrium 0.0338	0.0000	undef	0.0000
25	Uterus_Myometrium 0.0991	0.0340	2.9179	0.3427
	Uterus_allgemein 0.0509	0.1908	0.2669	3.7471
	Brust-Hyperplasie 0.0064			
	Prostata-Hyperplasie 0.0386			
	Samenblase 0.0801			
	Sinnesorgane 0.0588			
30	Weisse_Blutkoerperchen 0.0616			
	Zervix 0.1810			

	FOETUS
	%Haeufigkeit
35	Entwicklung 0.0139
	Gastrointestinal 0.0194
	Gehirn 0.0000
	Haematopoetisch 0.0275
40	Haut 0.0000
	Hepatisch 0.0000
	Herz-Blutgefuesse 0.0142
	Lunge 0.0145
	Nebenniere 0.0000
45	Niere 0.0247
	Placenta 0.0364
	Prostata 0.0499
	Sinnesorgane 0.0000

	NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN
	%Haeufigkeit
50	Brust 0.0068
	Eierstock_n 0.0000
	Eierstock_t 0.0101
55	Endokrines_Gewebe 0.0000
	Foetal 0.0064
	Gastrointestinal 0.0976
	Haematopoetisch 0.0057
60	Haut-Muskel 0.0259
	Hoden 0.0309
	Lunge 0.1802
	Nerven 0.0050
	Prostata 0.0274
	Sinnesorgane 0.0000
65	Uterus_n 0.0125

## 2.2 Fisher-Test

Um zu entscheiden, ob eine Partial-Sequenz  $S$  eines Gens in einer Bibliothek für Normal-Gewebe signifikant häufiger oder seltener vorkommt als in einer Bibliothek für entartetes Gewebe, wird Fishers Exakter Test, ein statistisches Standardverfahren (Hays, W. L., (1991) Statistics, Harcourt Brace College Publishers, Fort Worth), durchgeführt.

Die Null-Hypothese lautet: die beiden Bibliotheken können bezüglich der Häufigkeit zu  $S$  homologer Sequenzen nicht unterschieden werden. Falls die Null-Hypothese mit hinreichend hoher Sicherheit abgelehnt werden kann, wird das zu  $S$  gehörende Gen als interessanter Kandidat für ein Krebs-Gen akzeptiert, und es wird im nächsten Schritt versucht, eine Verlängerung seiner Sequenz zu erreichen.

## Beispiel 3

## Automatische Verlängerung der Partial-Sequenz

Die automatische Verlängerung der Partial-Sequenz  $S$  vollzieht sich in drei Schritten:

1. Ermittlung aller zu  $S$  homologen Sequenzen aus der Gesamtmenge der zur Verfügung stehenden Sequenzen mit Hilfe von BLAST
2. Assemblierung dieser Sequenzen mittels des Standardprogramms GAP4 (Bonfield, J. K., Smith, K. F., und Staden R. (1995), Nucleic Acids Research 23 4992-4999) (Contig-Bildung).
3. Berechnung einer Konsens-Sequenz  $C$  aus den assemblierten Sequenzen.

Die Konsens-Sequenz  $C$  wird im allgemeinen länger sein als die Ausgangssequenz  $s$ . Ihr elektronischer Northern-Blot wird demzufolge von dem für  $S$  abweichen. Ein erneuter Fisher-Test entscheidet, ob die Alternativ-Hypothese der Abweichung von einer gleichmäßigen Expression in beiden Bibliotheken aufrechterhalten werden kann. Ist dies der Fall, wird versucht,  $C$  in gleicher Weise wie  $S$  zu verlängern. Diese Iteration wird mit der jeweils erhaltenen Konsensus-Sequenzen  $C_i$  ( $i$ : Index der Iteration) fortgesetzt, bis die Alternativ-Hypothese verworfen wird (if  $H_0$  Exit; Abbruchkriterium I) oder bis keine automatische Verlängerung mehr möglich ist (while  $C_i > C_{i-1}$ ; Abbruchkriterium II).

Im Fall des Abbruchkriteriums II bekommt man mit der nach der letzten Iteration vorliegenden Konsens-Sequenz eine komplette oder annähernd komplette Sequenz eines Gens, das mit hoher statistischer Sicherheit mit Krebs in Zusammenhang gebracht werden kann.

Analog der oben beschriebenen Beispiele konnten die in der Tabelle I beschriebenen Nukleinsäure-Sequenzen aus Blasentumorgewebe gefunden werden.

Ferner konnten zu den einzelnen Nukleinsäure-Sequenzen die Peptidsequenzen (ORFs) bestimmt werden, die in der Tabelle II aufgelistet sind, wobei wenigen Nukleinsäure-Sequenzen kein Peptid zugeordnet werden kann und einigen Nukleinsäure-Sequenzen mehr als ein Peptid zugeordnet werden kann. Wie bereits oben erwähnt, sind sowohl die ermittelten Nukleinsäure-Sequenzen, als auch die den Nukleinsäure-Sequenzen zugeordneten Peptid-Sequenzen Gegenstand der vorliegenden Erfindung.

TABELLE I

Sequenz ID No.:	Wahrscheinlichkeit für eine spezifische Expression im Tumorgewebe %	Funktion	Länge des Ausgangs-EST in Basen	Länge der angemeldeten Sequenz in Basen	Chromosomale Lokalisation
1	99.54	Identisch zum humanen IgG aus V-D-J6 Region	202	202	
2	99.51	unbekannt	287	1926	unbekannt
3	99.2	unbekannt	196	762	unbekannt
4	98.58	unbekannt	295	918	unbekannt
5	97.43	unbekannt	303	1146	unbekannt
6	97.43	Identisch zum Kaposi Sarcoma-assoziierten Herpesvirus Glycoprotein M	173	2407	
7	95.29	Hyaluronectin (HN) ist ein Glykoprotein, dass Hyaluron bindet und oft in humanen Tumoren gefunden wird	204	1471	unbekannt
8	95.29	unbekannt	186	1732	unbekannt
9	95.29	unbekannt	197	989	unbekannt
10	91.23	unbekannt	150	150	unbekannt
11	91.23	unbekannt	286	1467	unbekannt
12	91.23	unbekannt	221	895	unbekannt
13	91.23	mitochondriales Enzym	223	467	Chromosom 7
14	91.23	Identisch zum humanen Collagen I (alpha-1 Kette)	198	511	
15	100	identisch zum humanen Keralin K7 (Typ II)	186	1899	
16	100	unbekannt	228	758	unbekannt
17	99.97	Identisch zum humanen Cofilin	123	302	
18	100	das H19 Gen wird nur vom maternalen Chromosom exprimiert und stellt möglicherweise ein Tumor-Supressorgen dar	230	824	unbekannt
19	100	Identisch zum humanen IGF-2 Wachstumsfaktor	80	2190	
20	99.56	unbekannt	271	2565	unbekannt
21	99.26	Identisch zum humanen IgV-L (Klon VL 29-1)	227	461	
22	100	Identisch zum humanen H19	213	2096	
23	98.77	mitochondriales Enzym	203	1348	unbekannt
24	100	Identisch zum humanen Anti-Hepatitis A IgG Variable Region	209	358	
25	96.49	Identisch zum humanen Saposin	89	89	
26	94.05	Gen, dass durch IL6 induziert wird	229	1632	unbekannt
27	94.05	unbekannt	195	2972	unbekannt
28	100	Identisch zu Immunoglobulin M schwere Kette V Region	207	496	
29	100	Identisch zur humanen Ig schweren Kette (variable Region)	212	397	
30	100	Identisch zur cDNA, die die leichte Kette eines monoklonalen Antikörpers kodiert, der gegen das humane Cytomegalovirus 65 kD Protein gerichtet ist	243	772	

Sequenz ID No.:	Wahrscheinlichkeit für eine spezifische Expression im Tumorgewebe %	Funktion	Länge des Ausgangs-EST in Basen	Länge der angemeldeten Sequenz in Basen	Chromosomale Lokalisation
31	90.26	unbekannt	210	1031	unbekannt
32	90.26	mitochondriales Enzym	184	739	unbekannt
33	100	Identisch zur Ig kappa leichten Kette variable Region D11.	89	651	
34	99.76	Identisch zum humanen hsp27	197	823	
35	100	Identisch zur Ig schweren Kettenvariablen Region V(2-1) (v(h)-iv Familie).	231	457	
36	93.95	unbekannt	203	1203	unbekannt
37	93.95	Identisch zur Immunoglobulin leichten Kette variable Region (lambda-Illb Untergruppe) von IgM Rheuma-bezogener Faktoren.	207	207	
38	100	Identisch zu US-Patent	135	346	
39	100	Identisch zu einer Immunoglobulin lambda leichten Kette (X57812)	237	926	
40	99.42	Identisch zur humanen hPGI mRNA, die das "bone small proteoglycan I" kodiert	73	2384	
41	99.99	Identisch zu patentierter Sequenz	194	334	
42	99.77	Identisch zu patentierter Sequenz	98	845	
43	99.11	Identisch zur CD24 "signal transducer mRNA"	178	2233	
44	97.85	Identisch zur humanen Ig schweren Kette variablen Region (Klon M49)	242	243	
45	94.61	unbekannt	220	817	unbekannt
46	100	Identisch zur Ig Alpha 1- Alpha 2m=Immunoglobulin A1-A2 lambda hybrid GAU schwere Kette {secreted alpha chain}	200	1644	
47	93.7	Identisch zu Pro-alpha 2(I) collagen (COL1A2) Gen	262	1133	
48	99.54	Identisch zu humanen NIC (Natural Inhibitor of Collagenase)	180	969	
49	99.28	Identisch zur humanen Ferritin L Kette	104	617	
50	97.9	Identisch zum humanen Calcyclin Gen (auch Prolactin-Rezeptor assoziiertes Protein)	210	704	

TABELLE II

5	DNA-Sequenzen Seq. ID. No.	Peptid-Sequenzen (ORF's) Seq. ID. No.
	2	128
		129
10		130
	3	131
		132
15		133
	4	134
		135
20		136
	5	137
		138
25		139
	7	140
		141
30		142
	8	143
	9	144
		145
35		146
	10	147
		148
40		149
	11	150
		151
45		152
	12	153
		154
50		155
	13	156
		157
55		158
	16	159
		160
60		161

65

DNA-Sequenzen Seq. ID. No.	Peptid-Sequenzen (ORF's) Seq. ID. No.	
18	162	5
	163	
	164	
	165	10
20	166	
	167	
	168	15
23	169	
	170	
26	171	20
	172	
27	173	
31	174	25
32	175	
	176	
	177	30
36	178	
	179	
	180	35
45	181	
	182	
	183	40

Die erfinderischen Nukleinsäure-Sequenzen Seq. ID No. 1 bis Seq. ID No. 50 der ermittelten Kandidatengene und die ermittelten Aminosäure-Sequenzen Seq. ID No. 51–106 werden in dem nachfolgenden Sequenzprotokoll beschrieben.

#### Sequenzprotokoll

### (1) ALLGEMEINE INFORMATION:

#### (i) ANMELDER:

(A) NAME: metaGen - Gesellschaft für Genomforschung mbH

(B) STRASSE: Ihnestrasse 63

(C) STADT: Berlin

(E) LAND: Deutschland

(F) POST CODE (ZIP): D-14195

(G) TELEFON: (030)-8413 1673

(H) TELEFAX: (030)-8413 1674

(ii) TITEL DER ERFINDUNG: Menschliche Nukleinsäure-Sequenzen aus  
Blasentumorgewebe

(iii) Anzahl der Sequenzen: 106

(iv) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk

(B) COMPUTER: IBM PC compatible

(C) OPERATING SYSTEM: PC-DOS/MS-DOS

(D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

(2) INFORMATION ÜBER SEQ ID NO: 1:

(i) SEQUENZ CHARAKTERISTIK:

(A) LÄNGE: 202 Basenpaare

(B) TYP: Nukleinsäure

(C) STRANG: einzel

(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung  
hergestellte partielle cDNA

(iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(C) ORGAN:

(vii) SONSTIGE HERKUNFT:

(A) BIBLIOTHEK: cDNA library

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 1

```
tgagagtcac ggacctcctg caaagaaca tgaacacact gtggttcttc ctctcctgg 60
tggcagctcc cagatgggtc ctgtcccagg tgcagctaca gcagtggggc gcaggactgt120
tgaagccttc ggagaccctg tcctaacat gcgtgtcttc cggtagactct tccagtactt180
actactggga ttgatccgc ca 202
```

(2) INFORMATION ÜBER SEQ ID NO: 2:

(i) SEQUENZ CHARAKTERISTIK:

(A) LÄNGE: 1926 Basenpaare

(B) TYP: Nukleinsäure

(C) STRANG: einzel

(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung  
hergestellte partielle cDNA

(iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(C) ORGAN:

(vii) SONSTIGE HERKUNFT:

(A) BIBLIOTHEK: cDNA library

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 2

```

ttgcgatggc t gatggactg tggctctcta accaaaggac cctagcgggc tcaacaattg 60
tcaagagcag ttggtgggtc tgaatacaat cctcagccaa ggatccctcc tgtgttacag 120
atggatcagc taaaacaagc caacactgaa gacacaaaga atgaggttag gttcattgaa 180
accaggggtaa cacctgtgga tgagctaaac acaaagatga caatgacctt gtaccaggta 240
tagaagctca gagacatgcc tgcaaaatga aatccctgag gaattttgca gctaccaga 300
gatacgtggg tcaaattaaa atgtctgacg gatcactcat ttgaggaaca gcacatcagc 360
ttcgcctctt acgtggacaa taggtttttt actttgacgg tgacaagtct ccacctggtg 420
ttccagatgg gagtcatatt cccacaataa gcagccctta ctaagccgag agatgtcatt 480
cctgcaggca ggacctatag gcacgtgaag atttgaatga aagtacagtt ccatttggaa 540
gccagacat aggatgggtc agtgggcatg gctctattcc tattctcaa ccatgccagt 600
ggcaacctgt gctcagtctg aagacaatgg acccagctta ggtgtgacac gttcacataa 660
ctgtgcagca catgccggga gtgatcagtc agacatttta atttgaacca cgtatctctg 720
ggtagctaca aaattcctca gggatttcat tttgcaggca tgtctctgag cttctatacc 780
tgctcaaggt cagtgtcatc tttgtgttta gctcatcaa aggtgttacc ctggtttcaa 840
tgaacctaac ctcatctctt gtgtcttcag tgttggcttg ttttagctga tccatctgta 900
acacaggagg gatccttggc tgaggattgt atttcagaac caccaactgc tcttgacaat 960
tgtaaacccg ctaggctctc ttggttagag aagccacagt ccttcagcct ccaattggtg1020
tcagtactta ggaagaccac agctagatgg acaaacagca ttgggaggcc ttagccctgc1080
tcctctcaat tccatcctgt agagaacagg agtcaggagc cgctggcagg agacagcatg1140
tcaccagga ctctgccggt gcagaatatg aacaatgcca tgttcttgca gaaaacgctt1200
agcctgagtt tcataggagg taatcaccag acaactgcag aatgtagaac actgagcagg1260
acaactgacc tgtctccttc acatagtcca tatcaccaca aatcacacaa caaaaaggag1320
aagagatatt ttgggttcaa aaaaagttaa aagataatgt agctgcattt ctttagttat1380
ttgaagccc caaatatttc ctcatctttt tgttgtgtc atggatggtg gtgacatgga1440
cttgtttata gaggacaggt cagctgtctg gctcagtgat ctacattctg aagttgtctg1500
aaaatgtctt catgattaaa ttcagcctaa acgttttgcc gggaacactg cagagacaat1560
gctgtgagtt tccaacctca gcccatctgc gggcagagaa ggtctagttt gtccatcacc1620
attatgatat caggactggt tacttggtta aggaggggtc taggagatct gtccctttta1680
gagacacctt acttataatg aagtacttgg gaaagtgtt ttcaagagta taaatacct1740
gtattctaat gatcatctc taaacatttt atcatttatt aatcctccct gcctgtgtct1800
attattatat tcatatctct acgctgcaa cttctgcct caatgtttac tgtgcctttg1860
tttttgctag tgtgtgttgt tgaaaaaaa aacattccct gcctaagtta gttttggcaal1920
agtatt

```

(2) INFORMATION ÜBER SEQ ID NO: 3:

(i) SEQUENZ CHARAKTERISTIK:

(A) LÄNGE: 762 Basenpaare

(B) TYP: Nukleinsäure

(C) STRANG: einzel

(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung hergestellte partielle cDNA



(iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(C) ORGAN:

(vii) SONSTIGE HERKUNFT:

(A) BIBLIOTHEK: cDNA library

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 3

```

ctccactgca accacccaga gccatggctc cccgaggctg catcgtagct gtctttgcca 60
ttttctgcat ctccaggctc ctctgctcac acggagcccc agtggccccc atgactcctt120
acctgatgct gtgccagcca cacaagagat gtggggacaa gttctacgac cccctgcagc180
actgttgcta tgatgatgcc gtcgtgccct tggccaggac ccagacgtgt ggaaactgca240
ccttcagagt ctgctttgag cagtgtgcc cctggacctt catggtgaag ctgataaacc300
agaactgcga ctcagcccg acctcggatg acaggctttg tcgcagtgtc agctaattga360
acatcagggg aacgatgact cctggattct ccttcctggg tgggcctgga gaaagaggct420
ggtgttacct gagatctggg atgctgagtg gctgtttggg ggccagagaa acacacactc480
aactgccac ttcattctgt gacctgtctg aggccaccc tgcgctgcc ctgaggaggc540
ccacagggcc ccttctagaa ttctggacag catgagatgc gtgtgctgat gggggcccag600
ggactctgaa cctcctgat gacctatg gccaacatca acccggcacc accccaaggc660
tggtgggga acccttcacc cttctgtgag atttccatc atctcaagtt ctcttctatc720
caggagcaaa gcacaggatc ataataaatt tatgtacttt aa 762

```

(2) INFORMATION ÜBER SEQ ID NO: 4:

(i) SEQUENZ CHARAKTERISTIK:

(A) LÄNGE: 918 Basenpaare

(B) TYP: Nukleinsäure

(C) STRANG: einzel

(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung hergestellte partielle cDNA

(iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(C) ORGAN:

(vii) SONSTIGE HERKUNFT:

(A) BIBLIOTHEK: cDNA library

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 4

```

ctcgagccgc tcgagccgat tcggctcgag ccttcccgtc cctgcttgc aaagtgggtg 60
tgccccaagg tccgcctcca ggccacgtgg gtgctgcggg ccaagctttc ccttcctttg120
agagaggttt ccgctgtagg agcagagctt ccgggctgcg ctcttcgttg cccagtttcc180

```

gctcagtggt cgcgtctccg cccccaccc accagtcacg ctgcattctc ggcggggctc240  
 taggcgccat ggctccccgc gggaggaagc gtaaggctga ggcgcggctg gtggccgtag360  
 ccgagaagcg agagaagctg gcgaacggcg gggagggaat ggaggaggcg accgttggtta360  
 tcgagcattg cactagctga cgcgtctatg ggcgcaacgc cgcggccctg agccaggcgc420  
 tgccgctgga ggccccagag cttccagtaa aggtgaaccc gacgaagccc cggaggggcag480  
 cttcgagggtg acgctgctgc gcccggaagg cagcagtgcg gaactctgga ctgggattaa540  
 gaagggggccc ccacgcaaac tcaaattccc tgagcctcaa gaggtggtgg aagagttgaa600  
 gaagtacctg tcgtaggag atttggtag aagccctcat gctgagcttt gtgtccctgg660  
 tgatgttgga acattaatga tggaaatgg ccaaacttca gtcattgatcc tgaagccatg720  
 gtttcttccc tgccagaaat gaaggttcag ttatgaggca accctctagt aaggcattgt780  
 aaaagttact ggatttggtt taataaaagt tgaaataaag taaaagaaaa aaaaaaaga840  
 gaaaaagaa aaagaaaaaa agaaagaaga aaaaagaaag gagaagcgag agaaaggag900  
 gccgcggggc gcggcggc 918

5

10

15

## (2) INFORMATION ÜBER SEQ ID NO: 5:

20

## (i) SEQUENZ CHARAKTERISTIK:

(A) LÄNGE: 1146 Basenpaare

(B) TYP: Nukleinsäure

(C) STRANG: einzel

(D) TOPOLOGIE: linear

25

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung  
hergestellte partielle cDNA

30

(iii) HYPOTHETISCH: NEIN

35

(iii) ANTI-SENSE: NEIN

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(C) ORGAN:

40

(vii) SONSTIGE HERKUNFT:

(A) BIBLIOTHEK: cDNA library

45

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 5

tcagtttagt ggcagggtgg ttttttaatt ttcttctgtg gctggatttt tgttgtttgt 60  
 ttaaataact cttctgggaa gttggtttat aagcctttgc cagggtgtaac tgttgtgaaa 120  
 taccacccac taaagttttt taagttccat attttctcca ttttgccctc ttatgtattt 180  
 tcaagattat tctgtgcaat ttaaatttac ttaacttacc ataatgacag tgtgactttt 240  
 cccacacact ggattgtgag gctcttaact tcttaaaagt ataatggcat cttgtgaatc 300  
 ctataagcag tctttatgtc tcttaacatt cacacctact ttttaaaaac aaatattatt 360  
 actattttta ttattgtttg tcctttataa attttcttaa agattaagaa aatttaagac 420  
 cccatttgag tactgtaatg caattcaact ttgagttatc ttttaaatat gtcttgata 480  
 gttcatattc atggctgaaa cttgaccaca ctattgctga ttgtatggtt ttacacctga 540  
 caccgtgtag aatgcttgat tacttgtagt cttcttatgc taatatgctc tgggctggag 600  
 aaatgaaatc ctcaagccat caggatttgc tatttaagtg gcttgacaac tgggccacca 660  
 aagaacttga acttcacctt ttaggatttg agctgttctg gaacacattg ctgcactttg 720  
 gaaagtcaaa atcaagtgcc agtggcgccc ttccataga gaatttgccc agctttgctt 780  
 taaaagatgt cttgtttttt atatacacat aatcaatagg tccaatctgc tctcaaggcc 840  
 ttggtcctgg tgggattcct tcaccaatta ctttaattaa aaatggctgc aactgtaaga 900  
 acccttgtct gatataattg caactatgct cccatttaca aatgtacctt ctaatgctca 960  
 gttgccaggt tccaatgcaa aggtggcgtg gactcccttt gtgtgggtgg ggtttgtggg1020  
 tagtggtgaa ggaccgatat cagaaaaatg ccttcaagtg tactaattta ttaataaaca1080

50

55

60

65

ttaggtgttt gttaaaaaaa aaaaaaaaaa aaaaaazag; ggtgggtcgt gttggtgtall140  
ttggtt 1146

5 (2) INFORMATION ÜBER SEQ ID NO: 6:

(i) SEQUENZ CHARAKTERISTIK:

- 10 (A) LÄNGE: 2407 Basenpaare  
(B) TYP: Nukleinsäure  
(C) STRANG: einzel  
(D) TOPOLOGIE: linear

- 15 (ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung  
hergestellte partielle cDNA

- 20 (iii) HYPOTHETISCH: NEIN

- (iii) ANTI-SENSE: NEIN

25 (vi) HERKUNFT:

- (A) ORGANISMUS: MENSCH  
(C) ORGAN:

30 (vii) SONSTIGE HERKUNFT:

- (A) BIBLIOTHEK: cDNA library

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 6

35  
gagtgaagtga gtgtgttgcg tcgaattaag gactcttgaa gagaagagag gtccattcag 60  
ggttgtccag attgaagtga ggtctcacgg tgaaaagaaa aggaaaatat tcagactctc 120  
ttgaaatcca aagagcaaga agtaaatgaa cttctgcaaa aattccagca agctcaggaa 180  
40 gaacttgtag aaatgaaaag atactctgag agctcttcaa aactggagga agataaagat 240  
aaaaagataa atgagatgtc gaaggaagtc accaaattga aggaggcctt gaacagcctc 300  
tcccagctct cctactcaac aagctcatcc aaaaggcaga gtcagcagct ggaggcgctg 360  
cagcagcaag tcaaacagct ccagaaccag ctggcggaat gcaagaaaca acaccaggag 420  
gtcatatcag tttaacagaat gcatcttctg tatgctgtgc agggccagat ggatgaagat 480  
45 gtccagaaag tactgaagca aatccttacc atgtgtaaaa accagtctca aaagaagtaa 540  
agtggattcc ttggcaggac actgcccctt gtcactctgc tttgtgttag atccagagtt 600  
gtggcagcc gctgccattg ttctcattcg tggatgcac tgtggcctag cgtagttctt 660  
ccctttccaa aggtttctga ggacttctcc caggagaaga ctgcccgcct cagaactgct 720  
tagagacttc aaaccagcag aggtgaaagt ccctgtcacc ccttcagatt ccagagctgg 780  
50 gatcagccat gcccagaggt ctggtcctga tgctggcagg ggggccccct cctccatccc 840  
tgactggctg agtggcttta tcaccaccga gtgatgtgct gaggcctcct gcagtgaatg 900  
ctccttccat tctgtactc gggcagtgcc attcagcaca ggagagctct ttttgccttt 960  
ggctttcaat tccaaaacat gatttaattt ctaactaaat tagtatggca ctagttatga1020  
55 agtatctgct taaaaccctt catcatgata tctgtggat ttaaaaactc taattccatg1080  
ttttcttccc atctgcctta tatactctcat caccctgctt atcaatatcc agtttgatga1140  
gcactattaa ctaaaatatg aaacttaaaa acaaaagcaa gttgtcctta aaagttcttt1200  
ttttaagtaa attgttgaca tactgcaaat ttctatgca aacttgccctc ctgctgttat1260  
ctgtgaagct caggaaatcc aaacatttgt gtttcaacaa gggacagtaa actgtgtgtt1320  
60 tacagccaaa agaaatgcct catagttctt aaacctcaact tttgtagaag tatttttttc1380  
tctgtaatat ttttattggc tcataaagat gttttcatat ctgaactcct aaataagtga1440  
aattacagta gattatatta acaaaatact ttttaggtag ccatgcttga gactttttta1500  
aaatataact ttttccttaa agttttcagc tatagcaaaa ggtagttatg tatgccagac1560  
ctaataatgag ctgccaccaa cacccttaga actttcagcc atgggtgtctt cagaattgtal620  
65 gcgcattttc gaacttagca atctctcctt ttaccggttg aatgttttga atgcctgac1680  
tctaccagcg ccataaatg atctctagaa ggactgttag taccaatctg tttttcaact1740  
ttgaagctaa aaaccctgat atggtaatat tatgggtgcat agcagaggtc tcggaaaaaal800  
aatattttctg ttcaactttac tttcaggtta aaaatgtttc taacacgctt gcaacttccc1860

ttatggcatt aatcttgttg agggagagag acagaatcct ggacctcca aatatttaq1923  
 ctgaaagtag ggcctgctct gacagggccc atgtccaca zggctjctty gcctcagtgg1930  
 gtgcttggt gtgctggatg atatgttgat ctgtattgga taaggaccaa tgacagcaaa2040  
 gcaaaaatgg ctttaaagct tgggtgttact tttcttaagt tgtttaatta tagttaagca2100  
 atttcaaaaa tgctccaaag aaatgtgaaa ggaccttttg tcacagcact tcagaaaata2160  
 cacaacagcc ccttctgccc ccgcacagaa atgctgcaga gtatataaaa cttgagacat2220  
 tttttagga tgctgacga ggtgtagcct tttatcttgt ttccggatgc atatttatta2280  
 cgagtactct ggtaaataat tgaaaagtta tatgctgtag ttttttagtat tttgtctttg2340  
 taatttacag aagttatttg agaaaataaa cttgtttcat tttgcaaaaa aaaaaaaaaa2400  
 atgaaaa 2407

## (2) INFORMATION ÜBER SEQ ID NO: 7:

## (i) SEQUENZ CHARAKTERISTIK:

- (A) LÄNGE: 1471 Basenpaare  
 (B) TYP: Nukleinsäure  
 (C) STRANG: einzel  
 (D) TOPOLOGIE: linear

## (ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung hergestellte partielle cDNA

## (iii) HYPOTHETISCH: NEIN

## (iii) ANTI-SENSE: NEIN

## (vi) HERKUNFT:

- (A) ORGANISMUS: MENSCH  
 (C) ORGAN:

## (vii) SONSTIGE HERKUNFT:

- (A) BIBLIOTHEK: cDNA library

## (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 7

ctgctgcaac ccggcggctc ctgcagcggg ggtcggctgt tgggtgtgga gtttcccagc 60  
 gccctcggg tccgaccctt tgagcgttct gctccggcgc cagctacctc gctcctcggc 120  
 gccatgacca caaccaccac cttcaaggga gtgcgaccca acagcaggaa tagctcccga 180  
 gttttcgggc ctccaggtgg tggatccaat ttttcattag gttttgatga accaacagaa 240  
 caacctgtga ggaagaacaa aatggcctct aatatctttg ggacacctga agaaaatcaa 300  
 gcttcttggg ccaagtcagc aggtgccaa gctagtgggt gcaggggaaga cttggagtca 360  
 tctggactgc agagaaggaa ctctctgtaa gcaagctccg gagacttctt agatctgaag 420  
 ggagaagggt atattcatga aaatgtggac acagacttgc caggcagcct ggggcagagt 480  
 gaagagaagc ccgtgcctgc tgcgcctgtg ccagcccgg tggccccggc ccagtgcca 540  
 tccagaagaa atccccctgg cggcaagtcc agcctcgtct tgggttagct ctgactgtcc 600  
 tgaacgctgt cgttctgtct gtttcttcca tgettgtgaa ctgcacaaact tgagcctgac 660  
 tgtacatctc ttggatttgb ttcattaaaa agaagcactt tatgtactgc tgtctttttt 720  
 tttttctttt tgaagaacag gtttctctct gtccttgact cttgggtctg tgggccatgg 780  
 catgagtgtt ttctagtgt agattggagg gaaagctttg tgacacttag tactgtgttt 840  
 ttaagaagaa ataatttgggt tccagatgtg ttagaggatc ttttgtactg aggtttttta 900  
 cactttactt gggtttacca agcctcaact ggacagacca taaacagtcc acaggcaccg 960  
 ttcttgccag gccccaacc acagggagtc tctccgcaga gccttcttgg tgttgcccta1020  
 acttgccagt ggcctttgct cagagcctcc tctgtgaca tgtgaacaaat gaagaggcct1080  
 gcgcctcctg ccttgccgcc tgcaaagcaa agaaactgcc ttttattttt taaccttaaa1140  
 aagtagccag atagtaacaa gactggctgg ctgatgagca aagcctttgc tctcacgcag1200  
 aggaaggctt ggatgtacaa tgaaactgcc ttggaactaa agcagtgaag caaggaggc1260  
 aatcacactg aagcgggtct tctccagga acgggggtccc acaggcgtgt tgttttaaat1320  
 aacctgatgc tgtgtgcatg atgctggtgc ttgaccatga aaggaaagtc tcatccttaa1380

aatgtgttgt acttcacaat cctggactgt tgcttcaagt aaacatatc cacattttgal440  
 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa a 1471

5 (2) INFORMATION ÜBER SEQ ID NO: 8:

(i) SEQUENZ CHARAKTERISTIK:

- 10 (A) LÄNGE: 1732 Basenpaare  
 (B) TYP: Nukleinsäure  
 (C) STRANG: einzel  
 (D) TOPOLOGIE: linear

- 15 (ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung  
 hergestellte partielle cDNA

- 20 (iii) HYPOTHETISCH: NEIN

- (iii) ANTI-SENSE: NEIN

25 (vi) HERKUNFT:

- (A) ORGANISMUS: MENSCH  
 (C) ORGAN:

30 (vii) SONSTIGE HERKUNFT:

- (A) BIBLIOTHEK: cDNA library

35 (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 8

gcagaaccta cgctgacgg gccggcgcc ggctgagccg cgctgcgcag cgacgcggga 60  
 atgaagcggg cgctgggcag gcgaaagggc gtgtggttgc gcctgaggaa gatacttttc 120  
 40 tgtgttttgg ggtgtacat tgccattcca tttctcatca aactatgtcc tggaatacag 180  
 gccaaactga ttttcttgaa tttcgtaaga gttccctatt tcattgattt gaaaaacca 240  
 caggatcaag gtttgaatca cacgtgtaac tactacctgc agccagagga agacgtgacc 300  
 attggagtct ggcacaccgt cctgcagtc tgggtggaaga acgccaagg caaagaccag 360  
 atgtggtatg aggatgcctt ggcttcacgc caccctatca ttctgtacct gcatgggaac 420  
 45 gcaggtacca gaggaggcga ccaccgcgtg gagctttaca aggtgctgag ttcccttggg 480  
 taccatgtgg tcacctttga ctacagaggt tggggtgact cagtgggaac gccatctgag 540  
 cggggcatga cctatgacgc actccacgtt tttgactgga tcaaagcaag aagtgggtgac 600  
 aaccccggtg acatctgggg ccactctctg ggcactggcg tggcgacaaa tctggtgcgg 660  
 cgctctgtg agcgaagac gcctccagat gcccttatat tggaaatctc attcactaat 720  
 50 atccgtgaag aagctaagag ccattccatt tcagtatat atcgatactt ccctgggttt 780  
 gactggttct tccttgatcc tattacaagt agtggaaatta aatttgcaaa tgatgaaaac 840  
 gtgaagcaca tctctgtcc cctgtctatc ctgcacgtg aggacgaccc ggtggtgccc 900  
 ttccagcttg gcagaaagct ctatagcatc gccgcaccag ctggaagctt ccgagatttc 960  
 55 aaagtccagt ttgtgccctt tcattcagac cttggctaca ggcacaaata cattacaag1020  
 agccctgagc tgccacggat actgagggaa ttccctgggg agtcggagcc tgagcaccag1080  
 cactgagcct ggccgtggga aggaagcatg aagacctctg ccctccctcc gtttccctcc1140  
 agtcagcagc ccggtatcct gaagcccgag ggggcgggca cctgcaatgc tcaggagccc1200  
 agctgcaccc tggagagcac ctacagatccc aggcggggag gccctgcag gcctgcagtg1260  
 60 cccggaggcc tgagcatggc tgtgtggaag gcgtgggtgg caggcatgtg gctctccttg1320  
 ccgcccctoa acctgagatc ttgttgggag acttaattggc agcaggcagc catcactgcc1380  
 tgggtgatgc tgcactgagc tggacagggg gaggccgggc aggggactct tggggctcgg1440  
 gacctatgct agctttttgg caccaccacac agagaacgtg ggggtccagg tctttctgca1500  
 65 ccttcccagc acatgcagaa tgactccagt ggttccatcg tccccctctg ccctgtgtac1560  
 ctgcttgctt ttctcagctg cccacacctcc cctgggctgg cccactcacc cacagtggaa1620  
 gtgcccggga tctgcacttc ctcccctttc acctacctgt acacctaacc tggccttaga1680  
 ctgagcttta ttttaagaata aaatcgtggt ggtggtcaaa aagacactct gc 1732

## (2) INFORMATION ÜBER SEQ ID NO: 9:

## (i) SEQUENZ CHARAKTERISTIK:

- (A) LÄNGE: 989 Basenpaare  
 (B) TYP: Nukleinsäure  
 (C) STRANG: einzel  
 (D) TOPOLOGIE: linear

## (ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung hergestellte partielle cDNA

## (iii) HYPOTHETISCH: NEIN

## (iii) ANTI-SENSE: NEIN

## (vi) HERKUNFT:

- (A) ORGANISMUS: MENSCH  
 (C) ORGAN:

## (vii) SONSTIGE HERKUNFT:

- (A) BIBLIOTHEK: cDNA library

## (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 9

```

cggtctgagc gtgatcgctg actcagctga ccctgcggga ccggaaaaag aaattcccgg 60
gccctggcct cttggcgcgca tgaggttccg gttctgtggt gatctggact gtcccgactg120
ggctcctggca gaaatcagca cgctggccaa gatgtcctct gtgaagtgc ggctgctctg180
caccaggtac taaaggagct gctgggacag gggattgatt atgagaagat cctgaagctc240
acggctgacg ccaagtttga gtcaggcgat gtgaaggcca cagtggcagt gctgagtttc300
atcctctcca gtgcggccaa gcacagtgtc gatggcgaat ccttgtccag tgaactgcag360
cagctggggc tgcccaaaga gcacgcggcc agcctgtgcc gctgttatga ggagaagcaa420
agccccctgc agaagcactt gcgggtctgc agcctacgca tgaatagggt ggcagggtgtg480
ggctggcggg tggactacac cctgagctcc agcctgctgc aatccgtgga agagcccatg540
gtgcacctgc ggctggagggt ggcagctgcc ccagggaccc cagcccagcc tgttgccatg600
tccctctcag cagacaagtt ccaggtcctc ctggcagaac tgaagcaggc ccagaccctg660
atgagctccc tgggctgagg agaagggtgt tccaggcctg tgtggagccg ccctgcccgt720
atggagtca cccctctgaa ctgctcttcg ggaggcagcc ctggttctag gatgctgagg780
ccctggccc gactctggcc tcccagatcc ccagctgcct cacttctctc ttgagaactt840
ggctcagggc tcctgaggac cttcccagc attaccttcc cttcccttga aaggcaattg900
ttggctgttt tcataagcag gaaaaataaa cagaagtata aaagagaaaa aaaaaaaaaa960
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa 989

```

## (2) INFORMATION ÜBER SEQ ID NO: 10:

## (i) SEQUENZ CHARAKTERISTIK:

- (A) LÄNGE: 150 Basenpaare  
 (B) TYP: Nukleinsäure  
 (C) STRANG: einzel  
 (D) TOPOLOGIE: linear

## (ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung hergestellte partielle cDNA

## (iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(C) ORGAN:

(vii) SONSTIGE HERKUNFT:

(A) BIBLIOTHEK: cDNA library

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 10

```

attttatgaa gttgaaaaat agctcacttt aaagctagtt ttgaagacgt gcagctgtga 60
cttgggtctg gttgggggtg ttgtgttttg agtcagccgt tttcactccc actgaggttg120
tcagaacatg cagattgctt cgattttctc                                     150

```

(2) INFORMATION ÜBER SEQ ID NO: 11:

(i) SEQUENZ CHARAKTERISTIK:

(A) LÄNGE: 1467 Basenpaare

(B) TYP: Nukleinsäure

(C) STRANG: einzel

(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung  
hergestellte partielle cDNA

(iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(C) ORGAN:

(vii) SONSTIGE HERKUNFT:

(A) BIBLIOTHEK: cDNA library

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 11

```

cgaggaccgg ccttgcgagc ggcgacgact ataaaatggc gcgtgctgca acccgcgccc 60
gcttcggaga gagaaatgct ggggtgcagc ttcaagctta ggaccacca ccatgcctat 120
ccaggtgctg aagggcctga ccatcactca ttaagaacag aggaggctgc ctgttactcc 180
tggtgttgca tccctccaga cactctgctg tttcctgcct aggcgtggct gcagccatgg 240
ctaggaaagc gctgccaccc acccacctgg gccagagctg gttctgctcc tgctgcaggg 300
aactgagct ggctatctcg gcgcttcggg caagaactgc aacaggctct cctgggtcct 360
gcaggtgtac agccggggcc ctgccttgct cctcagctct cgagagctgc tgctgccggg 420
tgacctgac caacctgata aggtgccatc ttcagctacc actgcaaggc cctgagggca 480
acagcagcac ggcactgccc acccggtctg tgatggcctg gtgccagctg ggagtcctcc 540
cggcacttcg aggcactga gccaccctc cagccccagc ccaccatgga cagggtatc 600
cagcttcctc ctcaacctcg tccctgccc ctgagccagt gacgcccaag gacatgcctg 660
ttaccacagg cctgtaccag cactagctgg tcaaggcat gacagtgtct gaggccgtct 720
tgagatcca ggccatcact ggcagcagc tgctctccat ggtgccaggg cccgccaggc 780
caccaggctc atgctgggac ccaaccagt gcacaaggac ttggctgctg agccacacac 840

```

ccaggagaag gtggataagt gggctaccaa gggcttcttg caggctaggg gaggagccac 900  
 ccccgcttcc ctattgtgac caggcctatg gggaggagct gtccatagc caccgtgaga 960  
 cctgggcctg gctctcaagg acagacaccg cctggcctgg tgctccaggg gtgaagcagg1020  
 ccagaatcct gggggagctg ctctgtggtt gagctgcatt caggaagtgc gggacatggt1080  
 aggggaggca aaaagccttg ggcactaccc tccctgtgga gctgttcggg gtccgtcgag1140  
 ctagccacac cctgacacca tgttcaaggg taccggaaga gaagggtgtc tgccccaac1200  
 ctccccctgt ggtgtcactg gccagatgtc atgagggaag caggccttgt gagtggacac1260  
 tgaccatgag tccctggggg gagtgatccc ccaggcatcg tgtgccatgt tgcacttctg1320  
 cccaggcagc aggggtgggtg ggtaccatgg gtgcccaccc ctccaccaca tggggcccca1380  
 aagcactgca ggccaagcag ggcaaccca cacccttgac ataaaagcat cttgaagctt1440  
 ttaaaaaaaa aaaaaaaa aaaataa 1467

## (2) INFORMATION ÜBER SEQ ID NO: 12:

## (i) SEQUENZ CHARAKTERISTIK:

(A) LÄNGE: 895 Basenpaare

(B) TYP: Nukleinsäure

(C) STRANG: einzel

(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung hergestellte partielle cDNA

(iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

## (vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(C) ORGAN:

## (vii) SONSTIGE HERKUNFT:

(A) BIBLIOTHEK: cDNA library

## (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 12

ctcgagccgc tcgagaacct cacttcttta ctctccaaa aagaagtggg gaaagaacca 60  
 tcaaaccttt cctcctgact taccaaacca ggaaaacagc aggagagggt ggctcaggac120  
 ttagggacag ggtatagctt agatgggtga aagcaaaagg gagcaggaag ttgtaaatca180  
 ctggctaatt agaaaaggag acagctaact ctaggatgaa gctgtgacta ggctggagtt240  
 gcttccttga agatgggact ccttgggtat caagacctat gccacatcac actggggcta300  
 ggggaagtgg tgatgccagc cctcaagtct gtcttcagcc agggacttga gaagttatat360  
 tgggcagtgg ctccaatctg tggaccagta ttccagcttt ccctgaagat caggcagggt420  
 gccattcatt gtctttctct cctagccccc tcaggaaaaga aggactatat ttgtactgta480  
 ccctaggggt tctggaaggg aaaacatgga atcaggatc tatagactga taggcctat540  
 ccacaagggc catgactggg aaaaggtatg ggagcagaag gagaattggg attttaggg600  
 gcagtacgct caccctaaa ttttgggtgg ctggggcatg tcttgaggcc cagactgtta660  
 agcaggctct gctggcctgt ttactcgtea ccacctctgc acctgctgtc ttgagactcc720  
 atccagcccc aggcacgcca cctgctcctg agcctccact atctccctgt gacgggtgaa780  
 cttegtgtac tgtgtctcgg gtccatatat gaattgtgag cagggttcat ctattttaaa840  
 cacagatgtt tacaaaataa agattatttc aaacccaaaa gaaaaaaaaa aaaaa 895

## (2) INFORMATION ÜBER SEQ ID NO: 13:

## (i) SEQUENZ CHARAKTERISTIK:

(A) LÄNGE: 467 Basenpaare



- (B) TYP: Nukleinsäure  
(C) STRANG: einzel  
(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung hergestellte partielle cDNA

(iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

- (vi) HERKUNFT:  
(A) ORGANISMUS: MENSCH  
(C) ORGAN:

(vii) SONSTIGE HERKUNFT:  
(A) BIBLIOTHEK: cDNA library

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 13

```

accagcagc cctcgcgcg tccggcacag cggacaccag gactccaaaa tggcgtcagt 60
tgtaccagtg aaggacaaga aacttctgga ggtcaaactg ggggagctgc caagctggat120
cttgatgcgg gacttcagtc ctagtggcat ttctggagcg tttcaaagag gttactaccg180
gtactacaac aagtacatca atgtgaagaa ggggagcatc tcgggggatta ccatgggtgct240
ggcatgctac gtgctcttta gctactcctt ttcttacaag catctcaagc acgagcggct300
ccgcaaatac cactgaagag gacacactct gcaccccccc accccacgac cttggcccga360
gcccctccgt gaggaacaca atctcaatcg ttgctgaatc ctttcatatc ctaataggaa420
ttaacctcca aataaaacat gactggtacg tgtaaaaaaa aaaaaaa 467
    
```

(2) INFORMATION ÜBER SEQ ID NO: 14:

- (i) SEQUENZ CHARAKTERISTIK:  
(A) LÄNGE: 511 Basenpaare  
(B) TYP: Nukleinsäure  
(C) STRANG: einzel  
(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung hergestellte partielle cDNA

(iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

- (vi) HERKUNFT:  
(A) ORGANISMUS: MENSCH  
(C) ORGAN:

(vii) SONSTIGE HERKUNFT:  
(A) BIBLIOTHEK: cDNA library

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 14

actagttcta gatcgcgagc ggccgccttt tttgttcaaa gtctatTTTT attctctttt 60  
 tctttttttt ttttttgtgg atggggactt gtgaatTTTT cttaaagggtgc tatttaacat120  
 gggaggagag cgtgtgcggc tccagccag cccgtgtctc actttccacc ctctctccac180  
 ctgcctctgg ctctctcaggc ctctgtctct cgacctctct cctctgaaac cctcctccac240  
 agctgcagcc catcctccc gcgcctcct agtctgtcct gcgtcctctg tccccgggtt300  
 tcagagacaa ctccccaaag cacaagcag tttttccccc taggggtggg aggaagcaaa360  
 agactctgta cctatTTTTgt atgtgtataa taatttgaga tgtttttaat tattttgatt420  
 gctggaataa agcatgtgga aatgaccca aacataaaaa aaaaaaaaaa aaaaaaaaaa480  
 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa a 511 10

## (2) INFORMATION ÜBER SEQ ID NO: 15:

## (i) SEQUENZ CHARAKTERISTIK:

(A) LÄNGE: 1899 Basenpaare

(B) TYP: Nukleinsäure

(C) STRANG: einzel

(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung  
hergestellte partielle cDNA

(iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

## (vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(C) ORGAN:

## (vii) SONSTIGE HERKUNFT:

(A) BIBLIOTHEK: cDNA library

## (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 15

tctccaccct ccccttcctt ctttcttttt ccccttcttg cactgggac actcaggcct 60  
 cagaatgagg ctgctttatt ggaagctatt ctgacatcac tttccagact gtctcactgt 120  
 cttgggacca ggcattggag gcaggggtgg gaatcttctt gtgattgtgg gtggtggctg 180  
 gaggagtgga gtggtgggag gcggctcagt cgcgggcact cctgcgactg gcggatgcgg 240  
 tccgatgga ataagccttc aggagcccag gaccgcact gctggagaag ctcagggcat 300  
 tgtgtcccat ggttcccccg aggttcagcc caatgccacc gccactgcta ctgccaccag 360  
 tggaattcat cacagagata ttcacggctc ccactccatc tccagccaac cggtctcct 420  
 cgccctccag cagcttgagg taggtggcga tctcgatgtc cagggccagc ttcacgctca 480  
 tgagtctctg gtactcacgc agctgccgtg ccatatctctg cttggcccgc tgcagggcgg 540  
 cttccagctc ctctgtctg gcaagagcat ccttgagcgc cagctcccca cgctcctcag 600  
 cctcggcaat ggcggcctcc aacttggcac gctggttctt gatgttgcg atctcagcct 660  
 cagcctctg gtagtggcgg ttcattctctg aaatctcatt ccgggtattc cggaggtcgt 720  
 ccccatgctt cccagcctgg gcctggaggg tctcaaaactt ggtctggtac caggcttcag 780  
 cctcagcccg gctgcatttg gccatctcct catactgcgc cttgacctca gcgatgatgc 840  
 cgtccaggtc cagggagcga ctgttgtcca tggacagcac cacagatgtg tcggagatct 900  
 gggactgcag ctctgtcaac tccgtctcat tgagggtcct gaggaagttg atctcatcat 960  
 tcagggcatc caccttgagg tccagctcca ccttgcctcat gtaggcagca tccacatcct1020  
 tcttcagcac cacaactca ttctcagcag ctgtgcggcg gtaatttca tcttcgtact1080  
 tattcttgaa gtctccacc acatcctgcc atctccgcag ctccgcctcc aggcggcccc1140  
 catccacctg cagtgcctca agctgacccc gaaggccagc aatctgggccc tcaaagatgt1200  
 ctgggaggcg gctgtctctg gccgacttct gctcctgcag cagcgtccac ttggtctcca1260

```

gcagcttggt ctgctgctcc agaaaccgca ccttytcgat gaagvaaggca aacttytgyt1320
tgagggtctt gatctgctcg ctctctctct ggcgäaccg 71ggagggag gggcggcgt1380
ccagccgcag cggggccagc aggctctggt taatggtgac ctgcgggatg ccggcggcca1440
5 ccgggcccc ataggcagag cgcacggcca cgcgggccc tgaggcgccg aggcgtaga1500
ggctgctgct gccaaggccg ccggggcgag cggagctcag gcgcacctgg gcgcggcggc1560
ccgagaaggc ggetgagcgc gaggtgaata ccggggagct gaagtggatg gacatggtgg1620
ctgggcccgg atggacctag cggcgggcca ggaggagcgc gactcgcctg acctcgggga1680
10 cactccgcac cttttatccg cgggagccgg tgctgggctt ccacaggtag gggcggggct1740
ggcgcggggc accgtttctc tgctgccagg cccctctgc gcgtccgtcc gccctctgcc1800
cgcccgcccc gccgaagccc aggttttcag tccaagcagg gatggtccgg agtaggcagg1860
agcgccatcc ctagacggcc gcagagaaca gcgggggac 1899

```

15 (2) INFORMATION ÜBER SEQ ID NO: 16:

(i) SEQUENZ CHARAKTERISTIK:

- 20 (A) LÄNGE: 758 Basenpaare  
 (B) TYP: Nukleinsäure  
 (C) STRANG: einzel  
 (D) TOPOLOGIE: linear

- 25 (ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung  
 hergestellte partielle cDNA

- 30 (iii) HYPOTHETISCH: NEIN

- (iii) ANTI-SENSE: NEIN

35 (vi) HERKUNFT:

- (A) ORGANISMUS: MENSCH  
 (C) ORGAN:

40 (vii) SONSTIGE HERKUNFT:

- (A) BIBLIOTHEK: cDNA library

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 16

45

```

cggtctgagc ggctcgagg gcaagaagaa catcacgtgc tgtgacaccg acttggtcaa 60
cgccagcggg gcccatgccc tgcagccggc tgccgccatc cttgcgctgc tccctgcact120
cggcctgctg ctctggggac cgggccagct ataggctctg gggggccccg ctgcagccca180
50 cactgggtgt ggtgccccag gcctctgtgc cactcctcac agacctggcc cagtgggagc240
ctgtcctggt tcctgaggca caccctaacg caagtctgac catgtatgtc tgcacccctg300
tgccccaccc tgaccctccc atggccctct ccaggactcc caccggcag atcagctcta360
gtgacacaga tccgctgca gatggccctt ccaacctct ctgctgctgt ttccatggcc420
cagcattctc cacccttaac cctgtgctca ggcacctctt ccccaggaa gccttccctg480
55 cccaccccat ctatgacttg agccaggctt ggtecggtgt gtcccccgca ccagcaggg540
gacaggcact caggagggcc cagtaaaggc tgagatgaag tggactgagt agaactggag600
gacaagagtc gacgtgagtt cctgggagtc tccagagatg gggcctggag gcctggagga660
agggggccagg cctcacatc gtggggctcc ctgaatggca gcctgagcac agcgtaggcc720
60 cttataaacc acctgttgga taagccaaaa aaaaaaaa 758

```

(2) INFORMATION ÜBER SEQ ID NO: 17:

65 (i) SEQUENZ CHARAKTERISTIK:

- (A) LÄNGE: 302 Basenpaare  
 (B) TYP: Nukleinsäure

(C) STRANG: einzel  
(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung  
hergestellte partielle cDNA

(iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

(vi) HERKUNFT:  
(A) ORGANISMUS: MENSCH  
(C) ORGAN:

(vii) SONSTIGE HERKUNFT:  
(A) BIBLIOTHEK: cDNA library

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 17

```
cggtctgacg gtctcgagat caagcatgaa ttgcaagcaa actgctacga ggaggtcaag 60
gaccgctgca ccttggcaga gaagctgggg ggcagtgccg tcctctccct ggagggaag120
cctttgtgag cccctttctg cgccttcttg cctgggagca tctgggcagg cccaacacc180
ttgccctttg ggggtttgca gggctcgccc cctttcctgg ccagaaccgg gaggggctg240
gggggggatt cccaggcagg gggggagggg ccaattccct tttcaacccc caggttgggc300
ca                                     302
```

(2) INFORMATION ÜBER SEQ ID NO: 18:

(i) SEQUENZ CHARAKTERISTIK:  
(A) LÄNGE: 824 Basenpaare  
(B) TYP: Nukleinsäure  
(C) STRANG: einzel  
(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung  
hergestellte partielle cDNA

(iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

(vi) HERKUNFT:  
(A) ORGANISMUS: MENSCH  
(C) ORGAN:

(vii) SONSTIGE HERKUNFT:  
(A) BIBLIOTHEK: cDNA library

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 18

```
ggcgaggaca gaggaggcgc gtccggcctt cctgaacacc ttaggctggt ggggctgagg 60
caagaagcgg gtctgtttct ttacttcctc cacggagtcg gcacactatg gctgccctg120
```

5 ggcctcccaga acccacaaca tgaaagaaat ggtggtaccc agctcaagnc tgggcctttg180  
 aatccggaca caaaaccctc tagcttggaa atgaatatgc tgcactttac aacctactgca249  
 ctacctgact caggaatcgg ctctggaagg tgaagctaga ggaaccagac ctcatcagcc300  
 caacatcaaa gacaccatcg gaacagcagc gcccgagca cccaccccg accggcgact360  
 ccatcttcat gggcaccctc tgcggcgagc ggttgaccac cagccaccac atcatcccag420  
 agctgagctc ctccagcggg atgacgcctt cccaccacc tccctcttct tctttttcat480  
 ccttctgtct ctttgtttct gagctttcct gtctttcctt ttttctgaga gattcaaagc540  
 ctccacgact ctgtttcccc cgtcccttct gaatttaatt tgcactaagt catttgact600  
 10 ggttgaggtt gtggagacgg ccttgagtct cagtacgagt gtgcgtgagt gtgagccacc660  
 ttggcaagtg cctgtgcagg gcccgccgc cctccatctg ggccgggtga ctggggcgcc720  
 ggctgtgtgc ccgaaggcct caccctggcc ctctggcctt agtctggga gggtccgaac780  
 cgaacatcaa gggaggcaag cttttcaagg catttcatt aatt 824

## (2) INFORMATION ÜBER SEQ ID NO: 19:

20

## (i) SEQUENZ CHARAKTERISTIK:

25

(A) LÄNGE: 2190 Basenpaare

(B) TYP: Nukleinsäure

(C) STRANG: einzel

(D) TOPOLOGIE: linear

30

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung hergestellte partielle cDNA

(iii) HYPOTHETISCH: NEIN

35

(iii) ANTI-SENSE: NEIN

40

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(C) ORGAN:

45

(vii) SONSTIGE HERKUNFT:

(A) BIBLIOTHEK: cDNA library

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 19

50 tttttttttt tttttttttt tttttttttt tttttttttt ttatcatgact tttaatgctt 60  
 tattgggatt gcaagcgtaa caagggtaaa gacaaaaccc aagcatggga ttttgccgga 120  
 aatattagcg ttaaaggagc tgagttgagt caaacacggg ccgcaagggt gaccgagggc 180  
 gcaggcacag gtgacattca gtgtttggcg tgggggtctt caggtgatgg cagaggaggg 240  
 gacccaagag ggggcccccc actgaagaca ttggggacac ggggaggaga caagatggag 300  
 55 agccacgact aggcacggag gtcagacagg cagcccgggc caggatggtt agtggcccag 360  
 gggagagctg caaacctggg gacgcaaggg gctggtcggc aagtgcctcc gggacacccc 420  
 actccggcga ggcagaatat aacactgggt ggggtgggtgt cctgacgaat gggcaggtaa 480  
 tttgggggtgc ctggaagcgt tttggatctc aggccaatgt gggttccaca attgtgacaa 540  
 tttggtctct tgggtctctg tccaatgttc cgaatggccc actcacaggg cgttgccga 600  
 60 gggaccctct gcgacgtcgc agctcgagcc gaaatgaggg aacccccaaa tttcatgtca 660  
 attgatctat tccccctctt tgtttcttgg ggcagttttt tttttacccc tccttagctt 720  
 tatgcgtcga gaaaccaaata taaaccccccc ccccatgtaa caggggggca gtgacaaaag 780  
 caagaacgca cgaagccagc ctggagacca ccacgtcctg ccccccgcca tttatcgccc 840  
 tgattggatt ttgtttttca tctgtccctg ttgcttgggt tgagttgagg gtggagcctc 900  
 65 ctgggggggca ctggccactg agcccccctg gagaagtcag aggggagtg agaaaggccac 960  
 tgtccggcct ggcttcttgg gacagtggct ggtccccaga agtcctgagg gcggaggggg 1020  
 gggttgggca ggtctcctc aggtgtcagg aggggtgctg gaggccacag gagggggctc 1080  
 ctggctggcc tgaggctggc cggaggggaa ggggctagca ggtgtgtaaa cagaggggttc 1140

catcaggctg	gggcagggtg	gccgccttcc	gcacacttga	ggaaccctcc	cctctccctc	120	
ggtgacatct	tgcccgcccc	tcagcaccct	gccttgcctc	caggagggtcc	gaagctctgt	1260	
gggacctctt	gggggcaagg	tggggtgagg	ccgggggagta	gggagggtcag	gcgggtctga	1320	
gccacagag	caggagagct	gccagggtctg	cccatcgacc	agggttgcttg	ggccccggag	1380	5
cccacgggtc	tggtgatgcc	atagcagcca	ccaccgcggc	gcctaggggt	gcggcaggga	1440	
ctcggcctct	gggagggtta	cctcgcccc	acttggtgcc	ccagctcagc	ccccctgcac	1500	
gcagccccgac	tagcagtcta	gaggcctgag	gcttctgggt	cctggtgacg	gggctggcat	1560	
gacccccggg	gtcgtccatg	ccagtccgcc	tcagtcgcag	agggtccctc	ggcaagcgcc	1620	
ctgtgagtgg	gccattcgga	acattggaca	gaagcccaaa	gagccaaatt	gtcacaattg	1680	10
tggaaccac	attggcctga	gatccaaaac	gcttcgaggc	accccaaatt	acctgcccat	1740	
tcgtcaggac	accacccac	ccagtgttat	attctgcctc	gccggagtgg	gtgttcccgg	1800	
gggcaacttg	cgaccagccc	cttgcgctcc	caggtttgca	gctctccctc	gggccactaa	1860	
ccatcctggc	ccgggtgcc	tgtctgacct	ccgtgcctag	tcgtggctct	ccatcttgct	1920	15
tcctcccgt	gtccccaatg	tcttcagtgg	ggggccccct	cttgggtccc	ctcctctgcc	1980	
atcacctgaa	gacccccacg	ccaaacactg	aatgtcacct	gtgectgccg	cctcggtcca	2040	
ccttgcgccc	cgtgtttgac	tcaactcagc	tcctttaacg	ctaataattc	cggcaaaatc	2100	
ccatgcttgg	gttttgtctt	taaccttgta	acgcttgcaa	tcaccaataa	gcattaaaag	2160	
tcatgaaatt	caaaaaaaaa	aaaaaaaaaa				2190	20

## (2) INFORMATION ÜBER SEQ ID NO: 20:

## (i) SEQUENZ CHARAKTERISTIK:

(A) LÄNGE: 2565 Basenpaare

(B) TYP: Nukleinsäure

(C) STRANG: einzel

(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung hergestellte partielle cDNA

(iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(C) ORGAN:

(vii) SONSTIGE HERKUNFT:

(A) BIBLIOTHEK: cDNA library

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 20

ctccccacc	tgtggccgc	aagccgtctg	tgggagtc	ggaccccgcc	ttccccagtt	60	
accctcgagc	tgagcccctt	actgctcctc	ccaccaatgg	gctccctcac	accaggaca	120	
ggactaagag	ggagctggcg	gagaatggag	gtgtcctgca	gctgggtggg	ccagaggaga	180	
agatgggctt	cccgggtcca	gactcacaga	aagagctggc	ctgaccacca	ggcacctcac	240	
tggcactgct	gacccatccc	agaacacaa	tctcagggac	ccgagcagct	ccaaggacga	300	
gaggatacag	cagacacaac	ctaatagaga	gggcgcctgc	agccttaacc	tcacggcct	360	60
tcgatactta	tgcaagcctg	gtgttgctcc	tgtcctcaga	gtcatcctgc	gtcatgctt	420	
tttcccgaat	gggttcacct	ctggcagttg	ccgcttcagt	cttgccctta	gcctcatctt	480	
gaagtgggta	gctggcggga	gaggtgggt	gcgccccctg	ctggccctga	ggctgcagag	540	
ttgggagcag	gacacctcac	ctgagtttca	ttttttttca	tgtccaaacc	atgcacatac	600	
tatagtccag	aatcaaagca	cttttgaaaa	gtggctgcat	ggccatcctc	cagggccag	660	65
gaagttgcat	tccaagggcc	tgtttacatg	gcagcagaat	ccatccccgg	cagtcagccc	720	
atagcttggg	accagctctg	gccctcctgc	ccagtcaggt	ttactcctct	tggttcctga	780	
aggtyggcaa	gtcattgtgt	tcccacaggc	ttctctaggg	tgggggcagg	tgtggggctg	840	

5 tggaattcca aagcaca aaa ggtgcagagg ggatrggctt tctgtgcct caactcacc 900  
 accaccctcc tgccttccag ttctgccagg tgcctcatgc tggggacaag taggagactg 960  
 ccaggggcccc aagaaatggg tgagcagtag agtcatctcg gggcacttgg cagtgtcaag 1020  
 10 cacttgcccc ttgcctcctt gaccacactg ggggtgggtg gccccagca cttcagaggc 1080  
 aggagccttt gggctgagca agcactgagg aggtggatgg aaggagagcat ctggaggggg 1140  
 ggagcttcct tgagcagtgg gcccaggcct ggccctccac acttcattct ctgacctttc 1200  
 tctctcctca ttccgggtgca tgtcctttct gcagctgcct ttcagcacag gtggttccac 1260  
 tgggggcagc taacgctgag tgacaaggat gggaaagccac aggtgcattt tactcaagtc 1320  
 15 ttctctagtc aatgaggggc acccagtgtc tctagggcag gctgggtggg ggtcccttag 1380  
 gtatcagcct ctcttactgt actctccggg aatgttaacc tttctatttt cagcctgtgc 1440  
 cactgtgcta ggaagctgg cttccccatt ggccctgtg ggtccacagc agcgtggctg 1500  
 cccccaggg ccaccgcttc tttcttgatc ctctttcctt aacagtgact tgggcttgag 1560  
 20 tctggcaagg aaccttgctt ttagcttcac caccaaggag agaggttgac atgacctccc 1620  
 cgccccctca ccaagctgg gaacagagg gatgtggtga gagccagggt cctctggccc 1680  
 tctccagggt gttttccact agtcaactg gtcttctcct tgtagcta atcaatcaat 1740  
 tcttcccttg cctgtgggca gtggagagtg ctgctgggtg tacgctgcac ctgcccactg 1800  
 agttggggaa agaggataat cagtgagcac tgttctgctc agagctcctg atctacccca 1860  
 25 ccccttagga tccaggactg ggtcaaagct gcataaaacc aggcctggc agcaacctgg 1920  
 gaatggctgg aggtgggaga gaacctgact tctctttccc tctccctcct ccaacattac 1980  
 tggaaactcta tctgttagg atcttctgag cttgtttccc tgctgggtgg gacagaggac 2040  
 aaaggagaag ggaggggtcta gaagaggcag ccttcttttg tctctgggg taaatgagct 2100  
 tgacctagag taaatggaga gacaaaagc ctctgatttt taatttccat aaaatgtag 2160  
 30 aagtatatat atacatatat atatttcttt aaatttttga gtctttgata tgtctaaaaa 2220  
 tccattccct ctgccctgaa gcctgagtga gacacatgaa gaaaactgtg tttcatttaa 2280  
 agatgttaat taaatgattg aaacttgaaa aaggctactg cttcttaatg ttggggggac 2340  
 agggcagtg tctgggccc ctttagaag ggaataatgt ttgcctgctg cacacattgg 2400  
 35 acccaagtat gggcctcttc tgcctagtac tgccaaaggg actgttaagg tgtcttgtcc 2460  
 atcttctacc cccaccccc cattacaggg taaagggaac cccagactag gtgagggggc 2520  
 agcagctgcc tcacacttgt gttctctcct gagatgggtc agctt 2565

35 (2) INFORMATION ÜBER SEQ ID NO: 21:

(i) SEQUENZ CHARAKTERISTIK:

- 40 (A) LÄNGE: 461 Basenpaare  
 (B) TYP: Nukleinsäure  
 (C) STRANG: einzel  
 (D) TOPOLOGIE: linear

- 45 (ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung hergestellte partielle cDNA

- 50 (iii) HYPOTHETISCH: NEIN

- (iii) ANTI-SENSE: NEIN

- 55 (vi) HERKUNFT:  
 (A) ORGANISMUS: MENSCH  
 (C) ORGAN:

- 60 (vii) SONSTIGE HERKUNFT:  
 (A) BIBLIOTHEK: cDNA library

- (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 21

65 gttcctccct agcggctcgc tcagaagcag agttctgggg tgtctccacc atggcctgga 60  
 ccctctctg gctcactctc ctactcttt gcatagggtc tgtggtttct tctgagctga 120  
 ctcaggaccc tgctgtgtct gtggccttgg gacagacagt caggatcaca tgccaaggag 180

acagcctcag aagctattat gcaagctggt accagcagaa gccaggagag gccctgtac240  
 ttgtcatcta tggtaaaaac aaccggccct cagggatccc agaccgattc tctggclcca300  
 gctcaggaaa cacagcttcc ttgaccatca ctggggctca ggcggaagat gaggctgact360  
 attacttgta aactccccgg gacagcaagt gggtaaccaa tgtgggtatt ccggcgggag420  
 ggaccaaacg ttgaccggtt cttaaggtca gcccaaaggg c 461

5

## (2) INFORMATION ÜBER SEQ ID NO: 22:

10

## (i) SEQUENZ CHARAKTERISTIK:

(A) LÄNGE: 2096 Basenpaare

(B) TYP: Nukleinsäure

(C) STRANG: einzel

15

(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung  
hergestellte partielle cDNA

20

(iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

25

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(C) ORGAN:

30

(vii) SONSTIGE HERKUNFT:

(A) BIBLIOTHEK: cDNA library

35

## (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 22

atccccgtca ctttttggtta caggacgtgg cagctggttg gacgagggga gctggtgggc 60  
 agggtttgat cccagggcct gggcaacgga ggtgtagctg gcagcagcgg gcaggtgagg 120  
 accccatctg ccgggcaggt gagtcccttc cctccccagg cctcgcttcc ccagccttct 180  
 gaaagaagga ggttttagggg atcgaggggt ggcggggaga agcagacacc ctcccagcag 240  
 aggggcagga tgggggcagg agagtttagca aaggtgacat cttctcgggg ggagccgaga 300  
 ctgcgcaagg ctgggggggtt atgggcccgt tccaggcaga aagagcaaga gggcagggag 360  
 ggagcacagg ggtggccagc gtaggggtcca gcagtgggg tggtacccca ggctgggtc 420  
 agacagggac atggcagggg acacaggaca gaggggtccc cagctgccac ctacccacc 480  
 gcaattcatt tagtagcagg cacaggggca gctccggcac ggctttctca ggcctatgcc 540  
 ggagcctcga gggctggaga gcgggaagac aggcagtgtc cggggagttg cagcaggacg 600  
 tcaccaggag ggcgaacggc cacgggaggg gggccccggg acattgcgca caaaggaggc 660  
 tgcaggggct cggcctgcgg gcgcccgtcc cagcaggcac tgcggcccag ggtctggtgc 720  
 ggagagggcc cacagtggac ttggtgacgc tgtatgccct caccgctcag cccctggggc 780  
 tggcttgga gacagtacag catccagggg agtcaagggc atggggcgag accagactag 840  
 gcgaggcggg cggggcgagg tgaatgagct ctcaggagg aggatgggtc aggcaggggt 900  
 gaggagcgca gggggcgggc agcgggaggg actggcctcc agagcccgtg gccaaggcgg 960  
 gcctcgcggg cgggcgacgga gccgggatcg gtgcctcagc gttcgggctg gagacgaggc 1020  
 caggtctcca gctggggtgg acgtgccac cagctgccga aggcaagacg ccaggtccgg 1080  
 tggacgtgac aagcaggaca tgacatggte cgggtgtgac gcgaggacag aggaggcgcg 1140  
 tccggccttc ctgaacacct taggctgggt gggctgcggc aagaagcggg tctgtttctt 1200  
 tacttctctc acggagtcgg cacactatgg ctgccctctg ggctcccaga acccacaaca 1260  
 tgaagagaaat ggtgctaccc agctcaagc tgggcctttg aatccggaca caaaacctc 1320  
 tagcttgga atgaatatgc tgcaactttac aacctgtca ctacctgact caggaatcgg 1380  
 ctctggaagg tgaagctaga ggaaccagac ctcatcagcc caacatcaaa gacaccatcg 1440  
 gaacagcagc gcccgagca cccacccgc accggcgact ccatcttcat ggccaccccc 1500  
 tgcggcgagc ggttgaccac cagccaccac atcatcccag agctgagctc ctccagcggg 1560  
 atgacgcgt cccaccacc tccctcttct tctttttcat ccttctgtct ctttgtttct 1620  
 gagctttcct gcttttcctt ttttctgaga gattcaaagc ctccacgact ctgtttcccc 1680

40

45

50

55

60

65



cgcccccttct gaatttaatt tgcactaagt catctccact ggttgagtt gtggagacgg1740  
 ccttgagttct cagtacgagt gtgcgtgagt gtgagccacc ttggcaagt cctgtgcagg1800  
 gcccggccgc cctccatctg ggccgggtga ctgggcgcgc gctgtgtgcc cgaggcctca1860  
 5 cctgtccctc gcctagtctg gaagctccga ccgacatcac ggagcagcct tcaagcattc1920  
 cattacgccc catctcgtc tgtgccctc cccaccaggg cttcagcagg agccctggac1980  
 tcatcatcaa taaacactgt tacagcaaaa aaaaaaaga aaaaaaaag aaaagaaaaa2040  
 aaagaaaaa aaaagaaaa aaagagagg aggaagaagg agaaaaggga gtgtgg 2096

10

## (2) INFORMATION ÜBER SEQ ID NO: 23:

15

## (i) SEQUENZ CHARAKTERISTIK:

(A) LÄNGE: 1348 Basenpaare

(B) TYP: Nukleinsäure

20

(C) STRANG: einzel

(D) TOPOLOGIE: linear

25

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung  
hergestellte partielle cDNA

(iii) HYPOTHETISCH: NEIN

30

(iii) ANTI-SENSE: NEIN

35

## (vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(C) ORGAN:

40

## (vii) SONSTIGE HERKUNFT:

(A) BIBLIOTHEK: cDNA library

## (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 23

45 ctctcgcgag gcccagaga gcaggcgtg ggcagtgtg aggtcgttg agtcacttcc 60  
 gcgtcaccag ctctgtgcc tgccagtcgg tgccctccc gctccagcca tgctctccgc 120  
 cctcgcccg cctgccagcg ctgctctccg ccgcagcttc agcacctcgg cccagaacaa 180  
 tgctaaagta gctgtgctag gggcctctgg aggcacgagg cagccacttt cacttctcct 240  
 gaagaacagc cccttggtga gccgcctgac cctctatgat atcgcgcaca caccggaggt 300  
 50 ggccgcagat ctgagccaca tcgagaccaa agccgctgtg aaaggctacc tcggacctga 360  
 acagctgcct gactgcctga aagggtgtga tgtggtagtt attccggctg gagtccccag 420  
 aaagccaggc atgacccggg acgacctgtt caacaccaat gccacgattg tggccaccct 480  
 gaccgtgcc tgtgcccagc actgcccga agccatgatc tgcgtcattg ccaatccggt 540  
 taattccacc atcccatca cagcagaagt tttcaagaag catggagtgt acaaccccaa 600  
 55 caaaatcttc ggcgtgacga ccctggacat cgtcagagcc aacacctttg ttgcagagct 660  
 gaagggtttg gatccagctc gactcaacgt ccctgtcatt ggtggccatg ctgggaagac 720  
 catcatcccc ctgatctctc agtgaccccc caaggtggac tttcccagg accagctgac 780  
 agcactcact gggcggatcc aggaggccgg caccggaggt gtcaaggcta aagccggagc 840  
 60 aggtctgcc accctctcca tggcgtatgc cggcgcgcgc tttgtcttct cccttgtgga 900  
 tgcaatgaat ggaaaggaag gtgtgtgga atgttccttc gtttaagtcac aggaaacgga 960  
 atgtacctac ttctccacac cgctgctgct tgggaaaaag ggcatcgaga agaacctggg1020  
 catcggcaaa gtctcctctt ttgaggagaa gatgatctcg gatgcatcc ccgagctgaa1080  
 ggcctccatc aagaagggg aagatttcgt gaagaccctg aagtgaaggc ctgtgacggg1140  
 65 tggccagttt ccttaattta tgaaggcatc atgtcactgc aaagccgttg cagataaaact1200  
 ttgtatttta atttgctttg gtgatgatta ctgtattgac atcatcatgc cttccaaatt1260  
 gtgggtggct ctgtgggcgc atcaataaaa gccgtccttg attttaaaaa aaaaaaaaaa1320  
 aaaaaaaaaa aaaaaaaaaa aaaaaaaa 1348

## (2) INFORMATION ÜBER SEQ ID NO: 24

## (i) SEQUENZ CHARAKTERISTIK:

- (A) LÄNGE: 358 Basenpaare  
 (B) TYP: Nukleinsäure  
 (C) STRANG: einzel  
 (D) TOPOLOGIE: linear

5

10

## (ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung hergestellte partielle cDNA

15

## (iii) HYPOTHETISCH: NEIN

## (iii) ANTI-SENSE: NEIN

20

## (vi) HERKUNFT:

- (A) ORGANISMUS: MENSCH  
 (C) ORGAN:

25

## (vii) SONSTIGE HERKUNFT:

- (A) BIBLIOTHEK: cDNA library

## (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 24

30

```

aatgggggac cggagaagaa gtacaagagc acgggtcagg ggcggggactc cgacggctcc 60
ttcttctctt acagcaggct aaccgtggac aagagcaggt ggcaggaggg gaatgtcttc120
tcattgtccg tgatgcatga ggctctgcac aaccactaca cgcagaagag cctctccctg180
tctccgggtä aatgagtgcg acggccggca agcccccgct ccccgggctc tcgcgggtgc240
acgaggatgc ttggcacgta ccccggtgac atacttccca ggcacccagc atggaataaa300
agcaccacgc gctgccctgg ggcccctgcg aaaaaaaga aaaagaatcg aaaagggg 358

```

35

40

## (2) INFORMATION ÜBER SEQ ID NO: 25:

## (i) SEQUENZ CHARAKTERISTIK:

- (A) LÄNGE: 89 Basenpaare  
 (B) TYP: Nukleinsäure  
 (C) STRANG: einzel  
 (D) TOPOLOGIE: linear

45

50

## (ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung hergestellte partielle cDNA

55

## (iii) HYPOTHETISCH: NEIN

## (iii) ANTI-SENSE: NEIN

60

## (vi) HERKUNFT:

- (A) ORGANISMUS: MENSCH  
 (C) ORGAN:

65

## (vii) SONSTIGE HERKUNFT:

(A) BIBLIOTHEK: cDNA library

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 25

gcccctagcc cctggcagac atagctgctt cagtgccctt tttcctctgc tggctagatg60  
 gatgttgatg cactggaggt acttttagc 89

(2) INFORMATION ÜBER SEQ ID NO: 26:

(i) SEQUENZ CHARAKTERISTIK:

- (A) LÄNGE: 1632 Basenpaare  
 (B) TYP: Nukleinsäure  
 (C) STRANG: einzel  
 (D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung  
 hergestellte partielle cDNA

(iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

(vi) HERKUNFT:

- (A) ORGANISMUS: MENSCH  
 (C) ORGAN:

(vii) SONSTIGE HERKUNFT:

(A) BIBLIOTHEK: cDNA library

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 26

gacactggtt ggttctgata agaggcaggg gaggagaaag ccgaggaaga gggagttgcg 60  
 gaagaggagg gaggtaacaa gttctcttat ccaccatcac accgggagtg ttgtccagcc 120  
 gtggaggagg aggacgatga agaagctgta aagaaagaag ctacacagac ctctacttct 180  
 gccttgtctc caggatccaa gccagcact tgggtgtctt gccagggga ggaagagaat 240  
 caagccacgg aggataaaag aacagaaaga agtaaaaggag ccaggaagac ctccgtgtcc 300  
 ccccgatctt caggctccga cccaggtcc tgggagtatc gttcaggaga ggcgtccgag 360  
 gagaaggagg aaaaggcaca caaagaaact gggaaaggag aagctgccc agggccgcaa 420  
 tcctcagccc cagcccagag gcccagctc aagtcctggt ggtgccaacc cagtgatgaa 480  
 gaggagggtg aggtcaaggc ttgggggca gctgagaagg atggagaagc tgagtgtcct 540  
 ccctgcatcc cccaccaag tgccttcctg aaggcctggg tgtattggcc aggagaggac 600  
 acagaggaag aggaagatga ggaagaagat gaggacagt actctggatc agatgaggaa 660  
 gagggagaag ctgaggcttc ctcttcact cctgctacag gtgtcttctt gaagtcctgg 720  
 gtctatcagc caggagagga cacacagtga tacaggatca gccaggatg aaagagaagc 780  
 tgagacttct gttccacac cccctgcaag tgctttcttg aaggcctggg tgtatcggcc 840  
 aggagaggac actggatagt gaggataagg aagatgattc agaagcagcc ttaggagaag 900  
 ctgagtcaga cccacatccc tcccaccgg accagagggc ccacttcagg ggctggggat 960  
 atcgacctgg aaaagagaca gaggaagagg aagctgctga ggactgggga gaagctgagc1020  
 cctgcccctt ccagtgggc atctatgtac ctggagagaa gccaccgct ccctgggctc1080  
 ctctaggtt gccctccga ctgcaaaggc ggctcaaggc ccagaaacc cctactcatg1140  
 atccggacct tgagactccc ctaaaaggcca gaaaggtgag cttctccgag aaggtcactg1200  
 tccatttctt ggctgtctgg gcaggggcgg cccaggccgc ccgcccaggc ccctgggagc1260  
 agcttgctcg ggatcgagc cgcttcgac gccgcatcac ccaggcccag gaggagctga1320  
 gccctgctt caccctgct gccggggcca gagctggggc acgctcagg aaccacctt1380  
 tagcccccct cctgcccct accagacct cctcttcctc ctctgtccct tcgtccccag1440  
 tccagaccac gcccttgagc caagctggg ccacacctc ccgctcgtc gctgctgcag1500  
 cggctgccct ggacctcagt gggaggcgtg gctgagacca actggtttgc ctataattta1560

ttaactatattt atttttttcta agtgtgggtt tatataagga ataaagcctt ttgatttgta1620  
acgaaaaaaa aa 1632

## (2) INFORMATION ÜBER SEQ ID NO: 27:

## (i) SEQUENZ CHARAKTERISTIK:

(A) LÄNGE: 2972 Basenpaare

(B) TYP: Nukleinsäure

(C) STRANG: einzel

(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung  
hergestellte partielle cDNA

(iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

## (vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(C) ORGAN:

## (vii) SONSTIGE HERKUNFT:

(A) BIBLIOTHEK: cDNA library.

## (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 27

ccaggacgag cacctcatta cattcttcgt gcctgtcttt gagccgctgc cccctcagta 60  
cttcacccga gtggtgtctg accgttggt ctcttgtgag acccagctgc ctgtctcctt 120  
ccggcacctg atcttgccgg agaagtaccc cctccaacc gaacttttgg acctgcagcc 180  
cttgcccgtg tctgtctga gaaacagtgc ctttgagagt ctttaccag ataaatttcc 240  
tttcttcaat cccatccaga cccaggtgtt taacactgta tacaacagtg acgacaacgt 300  
gtttgtgggg gccccacgg gcagcgggaa gactatttgt gcagagtgtg ccacctgcg 360  
aatgtgtctg cagagctcgg agggcgctg tgtgtacat acccccatgg aggcctggc 420  
agagcaggta tacatggact ggtacgagaa gttccaggac aggtcaaca agaaggtgg 480  
actcctgaca ggcgagacca gcacagacct gaagctgctg ggcaaaggga acattatcat 540  
cagcacccct gagaagtggg acatacttcc ccggcgatgg aagcagcgca agaactgca 600  
gaacatcaac ctcttcgtgg tggatgaggt ccacctatc gggggcgaga atgggctgt 660  
cttagaagtg atctgctccc gaatgcgcta catctctcc cagattgagc ggccattcg 720  
cattgtggca ctcagctctt cgctctccaa tgccaaggat gtggccact ggctgggctg 780  
cagtgccacc tccaccttca acttccatcc caatgtgcgt cccgtcccct tggagctgca 840  
catccagggc ttcaacatca gccatacaca aaccgcctg ctctccatgg ccaagcctgt 900  
gtaccatgat ataccaagc actcgcccaa gaagcctgtc attgtctttg tgccgtctcg 960  
caagcagacc cgcctcactg ccattgacat cctcaccacc tgtgcagcag acatccaacg1020  
gcagagggtc ttgactgca ccgagaagga tctgattccg tacctggaga agctaagtga1080  
cagcagctc aaggaaacgc tgctaaatgg ggtgggtac ctgcatgagg ggctcagccc1140  
catggagcga cgcctggtgg agcagctctt cagctcaggg gctatccagg tgggtggggc1200  
ttctcgaggt ctctgctggg gcatgaacgt ggtgcccac ctggtaatca tcatggatac1260  
ccagtactac aatggcaaga tccacgccta tgtgattac cccatctatg acgtgcttca1320  
gatggtgggc cagccaacc gccctttgca ggacgatgag gggcgctgtg tcatcatgtg1380  
tcagggtccc aagaaggatt tcttcaagaa gttcttatat gagccattgc cagtagaatc1440  
tcacctggac cactgtatgc atgaccactt caatgctgag atcgtcacca agaccattga1500  
gaacaagcag gatgctgtgg actacctcac ctggacctt ctgtaccgcc gcatgacaca1560  
gaaccceaat tactacaacc tgcagggcat ctcccatcgt cacttgtcgg accacttgtc1620  
agagctgggt gagcagacct tgagtgcctt ggagcagtc aagtgcata gcatcgagga1680  
cgagatggac gtggcgctc tgaacctagg catgatcgcc gcctactatt acatcaacta1740  
caccaccatt gagctcttca gcatgtccct caatgccaag accaaggtgc gagggttat1800

cgagatcatc tccaatgcag cagagtatga gaacattccc atccggcacc atgaagacaa1860  
 tctcctgagg cagttggctc agaaggtccc ccacazgctc aataacccta agltcaatga1920  
 tccgcacgtc aagaccaacc tgctcctgca ggctcacttg tctcgcatgc agctgagtgc1980  
 5 tgagttgcag tcagatacgg aggaaatcct tagtaaggca atccggctca tccaggcctg2040  
 cgtggatgtc ctttcagca atgggtggct cagccctgct ctggcagcta tggaaactggc2100  
 ccagatggtc acccaagcca tgtggtccaa ggactcatal ctgaagcagc tgccacactt2160  
 cactctgag catatcaaac gttgcacaga caagggagtg gagagtgtt tcgacatcat2220  
 ggagatggag gatgaagaac ggaacgcgtt gcttcagctg actgacagcc agattgcaga2280  
 10 tgtggctcgc ttttgaacc gctaccctaa tatcgaacta tcttatgagg tggtagataa2340  
 ggacagcatc cgcagtggcg ggccagtgtt ggtgctggtg cagctggagc gagaggagga2400  
 agtcacagge cctgtcattg cgcctctctt cccgcagaaa cgtgaagagg gctgggtggg2460  
 ggtgattgga gatgccaagt ccaatagcct catctccatc aagaggctga ccttgcagca2520  
 15 gaaggccaag gtgaagtgg actttgtggc cccagccact ggtgcccaca actacactct2580  
 gtacttcatt agtgacgctt acatgggatg tgaccaggag taaaaattca gcgtggatgt2640  
 gaaagaagct gagacagaca gtgattcaga ttgagtcctg aggcatttac ttttgggtaa2700  
 aggagagttg agcctgaatt aggaatgtgt acattgtagg aatcctggtt gtggggacca2760  
 ggtctgtggg cctcaggtct ggccagccag ggctggtgct gtccccgcct acctccactt2820  
 20 cctttccctt gctcactctg gatccagtga cagcaggtgt catgggtcaa gcataaatca2880  
 tatatagcat ttccagcat gttcctggta gttcttttga gtctgacatt ctaataaaa2940  
 aattttaga aaaaaacca aaaaaaaaaa aa 2972

25 (2) INFORMATION ÜBER SEQ ID NO: 28:

(i) SEQUENZ CHARAKTERISTIK:

- (A) LÄNGE: 496 Basenpaare  
 30 (B) TYP: Nukleinsäure  
 (C) STRANG: einzel  
 (D) TOPOLOGIE: linear

35 (ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung  
hergestellte partielle cDNA

40 (iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

45 (vi) HERKUNFT:

- (A) ORGANISMUS: MENSCH  
 (C) ORGAN:

50 (vii) SONSTIGE HERKUNFT:

- (A) BIBLIOTHEK: cDNA library

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 28

55 ctcgagccga agagtccctg acctcctgtg caagaacatg aaacatctgt ggttcttctt 60  
 tctcctgggt gcagctccca gatgggtcct gtcccagggt cagctgcagg agtcggggccc120  
 aggactgggt aagccttcgg agaccctgtc cctcacctgc actgtctctg gtggctccat180  
 60 cagtagttac tactggagct ggatccggca gcccccaggg aagggactgg agtggattgg240  
 gtatatctat tacagtggga gcaccaacta caaccctcc ctcaagagtc gattcaccat300  
 atcagtagac acgtccaaga accagttctc cctgaagctg agctctgtga ccgctgcgga360  
 cagggccgtg tattactgtg cgagacaggg tatagcagt gaccagcttg actactgggg420  
 ccagggaacc ctggtcaccg tctcctgagc ctgcaccaag gggccatcgg tcttccccct480  
 65 ggcaccctgc tccaag 496

(2) INFORMATION ÜBER SEQ ID NO: 29:

## (i) SEQUENZ CHARAKTERISTIK:

- (A) LÄNGE: 397 Basenpaare
- (B) TYP: Nukleinsäure
- (C) STRANG: einzel
- (D) TOPOLOGIE: linear

5

## (ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung hergestellte partielle cDNA

10

## (iii) HYPOTHETISCH: NEIN

15

## (iii) ANTI-SENSE: NEIN

## (vi) HERKUNFT:

- (A) ORGANISMUS: MENSCH
- (C) ORGAN:

20

## (vii) SONSTIGE HERKUNFT:

- (A) BIBLIOTHEK: cDNA library

25

## (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 29

```

gaggctcctgg acctcctgtg caagaacatg aaacacctgt ggttcttcct cctcctgggtg 60
gcagctccca gatgggtcct gtcccagggtg cagctgcagg agtcggggccc aggactgggtg120
aagccttcgg agacctgtc cctcacctgc actgtctctg gtggctccat cagtagttac180
tactggagct ggatccggca gcccgccggg aagggactgg agtggattgg gcgtatctat240
accagtggga gcaccaacta caaccctcc ctcaagagtc gagtcaccat gtcagtagac300
acgtccaaga accagttctc cctgaagctg agctctgtga ccgccgcgga cacggccgtg360
tattactgtg cgagagcaaa acgcagctgg acctcag 397

```

30

35

## (2) INFORMATION ÜBER SEQ ID NO: 30:

40

## (i) SEQUENZ CHARAKTERISTIK:

- (A) LÄNGE: 772 Basenpaare
- (B) TYP: Nukleinsäure
- (C) STRANG: einzel
- (D) TOPOLOGIE: linear

45

50

## (ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung hergestellte partielle cDNA

## (iii) HYPOTHETISCH: NEIN

55

## (iii) ANTI-SENSE: NEIN

## (vi) HERKUNFT:

- (A) ORGANISMUS: MENSCH
- (C) ORGAN:

60

## (vii) SONSTIGE HERKUNFT:

- (A) BIBLIOTHEK: cDNA library

65

## (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 30

```

5  gggactcctc agttcacctt ctcacaatga ggctccctgc tcagtcctg gggctgctaa 60
   tgctctgggt ctctggatcc agtggggata ttgtgatgac tcagtctcca ctctccctgc120
   ccgtcacccc tggagagccg gcctccatct cctgcaggtc tagtcagagc ctctgcata180
   gtaatggata caactatttg gattggtacc tgcagaagcc agggcagtc tccacagctcc240
   tgatctattt gggttctaata cgggcctccg gggccctga caggttcagt ggcagtggat300
10  caggcacaga ttttacactg aaaatcagca gaggggaggc tgaggatgtt ggggtttatt360
   actgcatgca agctctacaa actcctctca ctttcggcgg agggaccaag gtggagatca420
   aacgaactgt ggctgcacca tctgtcttca tcttcccgcc atctgatgag cagttgaaat480
   ctggaactgc ctctgttgtg tgctgtctga ataacttcta tcccagagag gccaaagtac540
   agtggaaggt ggataacgcc ctccaatcgg gtaactccca ggagagtgtc acagagcagg600
15  acagcaagga cagcacctac agcctcagca gcaccctgac gctgagcaaa gcagactacg660
   agaaacacaa agtctacgcc tgcgaagtca cccatcaggg cctgagcttg gcccgtttaag720
   aaagggcttt caacaggggg aagtttttag aggggagatg tggccccacc tt 772

```

20

## (2) INFORMATION ÜBER SEQ ID NO: 31:

## (i) SEQUENZ CHARAKTERISTIK:

- (A) LÄNGE: 1031 Basenpaare  
 (B) TYP: Nukleinsäure  
 (C) STRANG: einzel  
 (D) TOPOLOGIE: linear

30

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung hergestellte partielle cDNA

35

(iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

40

(vi) HERKUNFT:

- (A) ORGANISMUS: MENSCH  
 (C) ORGAN:

45

(vii) SONSTIGE HERKUNFT:

- (A) BIBLIOTHEK: cDNA library

50

## (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 31

```

55  ggggacggaa ggcagagca cggacccgc cccctcgcgg ccccgctcgt gacgtcgcgg 60
   ggggcgcggg cctccgccc gccccgagg gctctcccc gaggtcagc cccctctgct 120
   ccccatgggc aactgccagg cagggcacaa cctgcacctg tgtctggccc accaccacc 180
   tctggtctgt gccactttga tctgtctgct ccttggcctc tctggcctgg gccttggcag 240
   ctctcctcct acccacagga ctggcctgcg cagcctgaca tccccagga ctgggtctct 300
   tttttgagat cttttggcca gctgacctg tgtcccagga atgggacagt cacaggggag 360
60  tggcgagggt ctcacgtcgt gggcttgcgt accacctga acttcggaga cgggccagac 420
   aggaacaaga cccggacatt ccaggccaca gtcctgggaa gtcagatggg attgaaagga 480
   tcttctgcag gacaactggt ccttatcaca gccagggtga ccacagaaag gactgcagga 540
   acctgcctat attttagtgc tgttccagga atcctaccct ccagccagcc acccatatcc 600
   tgctcagagg agggggctgg aaatgccacc ctgagcccta gaatgggtga ggaatgtgtt 660
65  agtgtctgga gccatgaagg ccttggtgctg accaagctgc tcacctcgga ggagctggct 720
   ctgtgtggct ccaggctgct ggtcttgggc tccttctctg ttctcttctg tggccttctc 780
   tgctgtgtca ctgctatgtg cttccaccgc cgccgggagt cccactggtc tagaaccggg 840
   ctctgagggc actggcctag ttcccagact gtttctcagg tgtgaatcaa cttcttgggc 900

```

cttgctctg agttggaaaa ggttttagaa aaactgaaga gctggaatgt ggjggaaaat 960  
 aaaaagcttt ttgcccacaa aaaaaaaaaa aaaaaaaada aaaaaaaaaa aaaaaaaaaa1020  
 aaaaaaaaaa a 1031

5

## (2) INFORMATION ÜBER SEQ ID NO: 32:

## (i) SEQUENZ CHARAKTERISTIK:

(A) LÄNGE: 739 Basenpaare

10

(B) TYP: Nukleinsäure

(C) STRANG: einzel

(D) TOPOLOGIE: linear

15

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung  
hergestellte partielle cDNA

20

(iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

25

## (vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(C) ORGAN:

30

## (vii) SONSTIGE HERKUNFT:

(A) BIBLIOTHEK: cDNA library

## (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 32

35

cggctcgagc cccgctcagt caccgcgagc aggcgtgcag tttcccggt ctccgcgcgg 60  
 ccggggaagg tcagcgccgt aatggcggtt ttggcgctcg gaccctacct gaccatcag120  
 caaaaggtgt tgcggcttta taagcgggcg ctacgccacc tcgagtcgtg gtgcgtccag180  
 agagacaaat accgatactt tgcttggttg atgagagccc gggttgaaga acataagaat240  
 gaaaaggata tggcgaaggc caccagctg ctgaaggagg ccgaggaaga attctggtac300  
 cgtcagcatc cacagccata catcttccct gactctcctg gggcacctc ctatgagaga360  
 tacgattgct acaagggtcc agaattggtgc ttagatgact ggcatccttc tgagaaggca420  
 atgtatcctg attactttgc caagagagaa cagtgggaaga aactgcggag ggaaagctgg480  
 gaacgagagg ttaagcagct gcaggaggaa acgccacctg gtggtccttt aactgaagct540  
 ttgccccctg cccgaaagga aggtgatttg cccccactgt ggtggtatat tgtgaccaga600  
 ccccgggagc ggcccatgta gaaagagaga gacctcatct ttcattgctt caagtgaat660  
 atgttacaga acatgcactt gccctaataa aaaatcagtg aaatggaaaa aaaaaaaaaa720  
 aaaaaaaaaa aaaaaaaaaa 739

40

45

50

55

## (2) INFORMATION ÜBER SEQ ID NO: 33:

## (i) SEQUENZ CHARAKTERISTIK:

(A) LÄNGE: 651 Basenpaare

60

(B) TYP: Nukleinsäure

(C) STRANG: einzel

(D) TOPOLOGIE: linear

65

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung  
hergestellte partielle cDNA



(iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(C) ORGAN:

(vii) SONSTIGE HERKUNFT:

(A) BIBLIOTHEK: cDNA library

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 33

```

cggtctgagc ctcagttcac cttctcacca tgaggtctcc tgctcagctc ctggggctgc 60
taatgctctg ggtccctgga tccagtggag atattgtgat gaccagact ccactctccc120
tgccctgcac ccctggagag ccggcctcca tctcctgcag gtctagttag agcctcttg180
atagtgatga tggaaacacc tatttgact ggtacctgca gaagccagg cagtctccac240
agtcctgat ctatacgtt tctatcggg cctctggagt ccagacagg ttcagtggca300
gtgggtcagg cactgatttc acactgaaa tcagcagggt ggaggctgag gatgttgag360
tttattactg catgcaacgt atagaatttc cttacacttt tggccagggg accaagctgg420
agatcaaacg aactgtggct gcaccatctg tcttcattct cccgccatct ggatgagcag480
ttgaaatctg gaacttgctt ctgtgtgtgt gcctgcttga ataactttct attcccagag540
aggggcaaag taacagtggg aggttggtt aacgccctgc aattcgggta actgccagg600
gagtagtttt cacagggcag gggcagcaag gacagcacct acagtcttag t 651

```

(2) INFORMATION ÜBER SEQ ID NO: 34:

(i) SEQUENZ CHARAKTERISTIK:

(A) LÄNGE: 823 Basenpaare

(B) TYP: Nukleinsäure

(C) STRANG: einzel

(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung hergestellte partielle cDNA

(iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(C) ORGAN:

(vii) SONSTIGE HERKUNFT:

(A) BIBLIOTHEK: cDNA library

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 34

```

ccgcgtcgac aaattttttt aaagatcatc gatgaagaga gaaaatgcgc ttttctacag 60
agtcccttcc ccacccacag ccccatcccc agataagcgg ggagttccct ggcgcggtgc120
cagtttctag ccgctgagtg ggcgtgtgcg cggtcccaag tgcgcctgcg tactgctcac180
tccccagctc cgcgccttgc tccgttcctc ccaaaactct gaatcgaaga actttccgga240
agtttctgag agcccagacc ggcgggcacg cgcccatccc caacccctc tgttaatccc300
taccagcctg cagtcctggc tgcttccaag caggaggtgg ggctctggc ctacgggggc360

```

cgaaaggcag tgccctccc ccgcagtctg atttccctct tccccccaac ggcaagcacg120  
 aggagcggca ggacgagcat ggctacatct cccggtgctt cacgcggaaa tacacgtgc480  
 cccccggtgt ggaccccacc caagtttctt cctccctgtc cctgaggggc aactgaccg540  
 tggaggcccc catgcccaag ctagccacgc agtccaacga gatcaccatc ccagtcacct600  
 tcgagtcgcy ggcccagctt gggggcccag aagctgcaaa atccgatgag actgccgcca660  
 agtaaagcct tagcccggat gccaccccct gctgcgcgca ctggtgtgtc ctcccccgcc720  
 acctgtgtgt tcttttgata catttatctt ctgtttttct caaataaagt tcaaagcaac780  
 cacctggtca aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaa 823

5

10

## (2) INFORMATION ÜBER SEQ ID NO: 35:

## (i) SEQUENZ CHARAKTERISTIK:

15

(A) LÄNGE: 457 Basenpaare

(B) TYP: Nukleinsäure

(C) STRANG: einzel

20

(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung  
hergestellte partielle cDNA

25

(iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

30

## (vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(C) ORGAN:

35

## (vii) SONSTIGE HERKUNFT:

(A) BIBLIOTHEK: cDNA library

40

## (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 35

cataagggaa atgctttctg agagtcattg atctcatgtg caagaaaatg aagcacctgt 60  
 ggttcttctt cctgctgggtg gcggtcccca gatgggtcct gtcccagctg cagctgcagg120  
 agtcggggccc aggactgggtg aagccttcgg agaccctgtc cctcacctgc actgtctctg180  
 gtggctccat cagcagtagt agttactact ggggctggat ccgccagccc ccagggaagg240  
 ggctggagtg gattgggagt atctattata gtgggagcac ctactacaac ccgtccctca300  
 agagtcgagt caccatatcc gtagacacgt ccaagaagta cttctccctg aagctgagct360  
 ctgtgaccgc cgcagacacg gctgtgtatt actgtgcgag acatgactgg tattacgata420  
 ttttgactgg ttatgcgaaa cccggcacag gttcgac 457

45

50

## (2) INFORMATION ÜBER SEQ ID NO: 36:

55

## (i) SEQUENZ CHARAKTERISTIK:

(A) LÄNGE: 1203 Basenpaare

(B) TYP: Nukleinsäure

(C) STRANG: einzel

60

(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung  
hergestellte partielle cDNA

65

(iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(C) ORGAN:

(vii) SONSTIGE HERKUNFT:

(A) BIBLIOTHEK: cDNA library

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 36

```

15      gtcggggggcg cctgcgcagt cgctcttcct caggcggcgg ccatggcggg acaggaggat 60
      ccggtgcagc gggagattca ccaggactgg gctaaccggg agtacattga gataatcacc 120
      agcagcatca agaaaatcgc agactttctc aactcgttcg atatgtcttg tcgttcaaga 180
20      cttgcaacac taaacgagaa attgacagcc cttgaacgga gaatagagta cattgaagct 240
      cgggtgacaa aaggtagagc actcacctag aacagtgccg tgctgctgct gggaaagttgc 300
      ttacacaaac acaggccaca tgggaaaggc cccagcagcc ttcagctcct tcctttctcc 360
      ttaaagagca acagggctta ttcttggttt tcttttttca aaagtgtggc ctttgggctc 420
      tgccatctgg ggtgtggtgt ggtatgtggg aagaagttca gaggaaccgt tggaaacgac 480
25      tttaggcatt ttaccctttc agtaacattt tatacatcta cttgtcaatg tatttgagac 540
      attcacagcc aaaagcctgg gactctttgt gaaggctctc ctcacctcta tctttctttc 600
      tctctctctc aaactttcct taaagttctc attgccttg cactgcttct gtgaacagtc 660
      tttgtctcct cccaccttt ggtgggaagt gcggggcagt cctgggtcaag acactcatgc 720
30      cctggcaatg tggctgccag agaatgttgt tgctaaccga ccagtttctt gttgatttgg 780
      agaggtaag gccaggcccc cacttggttt gaaggacat tttcagactt ttctttctgt 840
      cacttgaggt gtctatgcct ctcatatttc cctaataaac tcctcaactt tttatctgac 900
      tgctgtgatt atggtgggga gaggagctag agatgggttc acttattgca cagaaatgta 960
      atacatggcg ttattattct aacataaaac tttcagatgt agctgtttga ttcaaagcct1020
35      aggtgcttac cagcccaagt ccccatgttt ggactttcag ctgactagct catcttggga1080
      atcatttgggt cattcagcac atttaccaag tatttactat gtaggcattg taaactccaa1140
      taaaacatac agcattgaat cagaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa1200
      aaa
  
```

(2) INFORMATION ÜBER SEQ ID NO: 37:

(i) SEQUENZ CHARAKTERISTIK:

(A) LÄNGE: 207 Basenpaare

(B) TYP: Nukleinsäure

(C) STRANG: einzel

(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung  
hergestellte partielle cDNA

(iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(C) ORGAN:

(vii) SONSTIGE HERKUNFT:

(A) BIBLIOTHEK: cDNA library

## (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 37

cggtctgagg ccgccctcgg tgtcagtgtc cccaggacag acggccagga tcacctgctc 60 5  
 tggacatgca ttgccaaagc aatgatgctt attggtacca gtcagaggcc agggccaggc120  
 ccctgtgctt ggtggatccc ttgaaagaac attggaggag ggcccttcag ggcattgcct180  
 ggagacggat tgctctgggc ttccaac 207

## (2) INFORMATION ÜBER SEQ ID NO: 38:

## (i) SEQUENZ CHARAKTERISTIK:

(A) LÄNGE: 346 Basenpaare

(B) TYP: Nukleinsäure

(C) STRANG: einzel

(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung  
 hergestellte partielle cDNA

(iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

## (vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(C) ORGAN:

## (vii) SONSTIGE HERKUNFT:

(A) BIBLIOTHEK: cDNA library

## (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 38

cggtctgagc ggctcgaggc cgggatggtg ggtgctacgc cccttgggta ctggggccag 60  
 ggaaccttg taccgtctc ctcagcctcc accaagggcc catcgggctt cccccgggca120  
 ccctcctcca agagcacctc tgggggcaca gcggccctgg gctgcctggt caaggactac180  
 ttccccgaac cggtgacggg gtcgtggaac tcaggcgcct gaccagcggc gtgcacacct240  
 tcccggctgt ctacagtctc aggactctac tcctcagcag cgtgggtgac tgccctccag300  
 cagttgggca ccagacctac atctgcaagt gaatcgaagc cagcaa 346

## (2) INFORMATION ÜBER SEQ ID NO: 39:

## (i) SEQUENZ CHARAKTERISTIK:

(A) LÄNGE: 926 Basenpaare

(B) TYP: Nukleinsäure

(C) STRANG: einzel

(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung  
 hergestellte partielle cDNA

(iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(C) ORGAN:

(vii) SONSTIGE HERKUNFT:

(A) BIBLIOTHEK: cDNA library

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 39

```

15  cggctctaag gaagcagcac tgggtggtgcc tcagccatgg cctggaccgt tctcctcctc 60
    ggccctcctc ctactgcac aggtctctgtg acctcctatg tgctgactca gccaccctcg120
    gtgtcagtg cccaggaca gacggccagg attacctgtg ggggaaacaa cattggaagt180
20  aaaagtgtgc actggtacca gcagaagcca ggccaggccc ctgtgctggt cgtctatgat240
    gatagcgacc ggccctcagg gatccctgag cgattctctg gctccaaactc tgggaacacg300
    gccaccctga ccatcagcag ggtcgaagcc ggggatgagg ccgactatta ctgtcaggtg360
    tgggatatga gtagtgatca ttgggtgttc ggcggaggga ccaagctgac cgtcctaggt420
    cagcccaagg ctgccccctc ggtcactctg ttcccgcctt cctctgagga gcttcaagcc480
25  aacaaggcca cactggtgtg tctcataagt gacttctacc cgggagccgt gacagtggcc540
    tggaaggcag atagcagccc cgtcaaggcg ggagtggaga ccaccacacc ctccaaacaa600
    agcaacaaca agtacgcggc cagcagctat ctgagcctga cgcctgagca gtggaagtcc660
    cacagaagct acagctgccg ggtcacgcat gaaggagca ccgtggagaa gacagtggcc720
    cctacagaat gttcataggt tctcaacct cccccccac caggggagac tagagtgcga780
30  ggatcccagg ggaggggtct ctctcccac cccaaggcat caagccctt tccctgcact840
    caataaacc tcaataaata ttctcattgt caatcaggaa aaaaaaaaaa aaaaaaaaaa900
    aaaaaaaaaa aaaaaaaaaa aaaaaa 926

```

(2) INFORMATION ÜBER SEQ ID NO: 40:

(i) SEQUENZ CHARAKTERISTIK:

(A) LÄNGE: 2384 Basenpaare

(B) TYP: Nukleinsäure

(C) STRANG: einzel

(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung hergestellte partielle cDNA

(iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(C) ORGAN:

(vii) SONSTIGE HERKUNFT:

(A) BIBLIOTHEK: cDNA library

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 40

```

65  gcctccgcc cgcgcctct gtctcctct ctccacaaac tgcccaggag tgagttagctg 60
    ctttcggtcc gccggacaca ccggacagat agacgtgcgg acggcccacc accccagccc 120
    gccaaactag cagcctgcgc ctggcgctc ccctctccag gtccatccgc catgtggccc 180

```

ctgtggcgcc	tgtgtgtctct	gctggccctg	agccaggccc	tgccctttga	gr:agagagggc	240	
ttctgggact	tcacctgga	cgatgggcca	ttcatgatga	acgatgagga	agcttcgggc	300	
gctgacacct	cgggcgctcct	ggaccgggac	tctgtcacac	ccacctacag	cgccatgtgt	360	
cctttcggt	gccactgcca	cctgcgggtg	gttcagtget	ccgacctggg	tctgaagtct	420	5
gtgccc aaag	agatctcccc	tgacaccacg	ctgctggacc	tgagaacaa	cgacatctcc	480	
gagctccgca	aggatgactt	caagggtctc	cagcacctct	acgccctcgt	cctggtgaac	540	
aacaagatct	ccaagatcca	tgagaaggcc	ttcagcccac	tgcggaagct	gcagaagctc	600	
tacatctcca	agaaccacct	ggtggagatc	ccgccc aacc	taccagctc	cctggtggag	660	
ctccgcatcc	acgacaaccg	catccgcaag	gtgccc aagg	gagtgttcag	tgggctccgg	720	10
aacatgaact	gcacgagat	gggcggaac	ccactggaga	acagtggctt	tgaacctgga	780	
gccttcgatg	gcctgaagct	caactacctg	cgcacctcag	aggccaagct	gactggcatc	840	
cccaaagacc	tccctgagac	cctgaatgaa	ctccacctag	accacaacaa	aatccaggcc	900	
atcgaaactg	aggacctgct	tgcctactcc	aagctgtaca	ggctgggctt	aggccacaac	960	15
cagatcagga	tgatcgagaa	cgggagcctg	agcttctctg	ccacctccg	ggagctccac	1020	
ttggacaaca	aaaagtggc	cagggtgccc	tcagggtccc	cagacctcaa	gctcctccag	1080	
gtggtctatc	tgactccaa	caacatcacc	aaagtgggtg	tcaacgactt	ctgtcccatg	1140	
ggcttcgggg	tgaagcgggc	ctactacaac	ggcatcagcc	tcttcaacaa	ccccgtgccc	1200	
tactgggagg	tgacgccggc	cactttccgc	tgctgactg	accgcctggc	catcpagttt	1260	20
ggcaactaca	aaaagtagag	gcagctgcag	ccaccgcggg	gcctcagtg	gggtctctgg	1320	
ggaacacagc	cagacatcct	gatggggagg	cagagccagg	aagctaagcc	agggcccagc	1380	
tgcgtccaac	ccagccccc	acctcggtc	cctgacccca	gctcgatgcc	ccatcaccgc	1440	
ctctccttg	ctcccaagg	tgaggtggg	cgcaaggccc	ggcccccatc	acatgttccc	1500	
ttggcctcag	agctgcccct	gctctccac	cacagccacc	cagaggcacc	ccatgaagct	1560	25
ttttctcgt	tactcccaa	acccaagtgt	ccaaggctcc	agtcctagga	gaacagtccc	1620	
tgggtcagca	gccaggaggc	ggtccataag	aatggggaca	gtgggctctg	ccagggtcgc	1680	
cgcacctgtc	cagacacaca	tgttctgttc	ctcctcctca	tgcatttcca	gcctttcaac	1740	
cctccccgac	tctgcggctc	ccctcagccc	ccttgcaagt	tcattggcctg	tccctcccag	1800	30
accctgtctc	cactggccct	tcgaccagtc	ctccctctg	ttctctcttt	ccccgtcctt	1860	
cctctctctc	tctctctctc	tttctgtgtg	tgtgtgtgtg	tgtgtgtgtg	tgtgtgtgtg	1920	
tgtgtgtgtc	ttgtgttcc	tcagaccttt	ctcgtctctg	agcttggtgg	cctgttccct	1980	
ccatctctcc	gaacctggct	tcgctgtcc	ctttcactcc	acacctctg	gccttctgcc	2040	
ttgagctggg	actgtttct	gtctgtccgg	cctgcaccca	gccccgtccc	acaaaacccc	2100	35
agggacagcg	gtctcccag	cctgcctgc	tcaggccttg	cccccaaac	tgtactgtcc	2160	
cggaggaggt	tgggaggtgg	agggccagca	tcccgcgag	atgacacat	caaccgccag	2220	
agtcacagac	accggttttc	ctagaagccc	ctcaccacca	ctggccact	ggtggctagg	2280	
tctccctta	tccttctggt	ccagcgcaag	gaggggctgc	ttctgaggtc	ggtggctgtc	2340	
tttcattaa	agaaacaccg	tgcaactgtg	aaaaaaaa	aaaa		2384	40

## (2) INFORMATION ÜBER SEQ ID NO: 41:

## (i) SEQUENZ CHARAKTERISTIK:

(A) LÄNGE: 334 Basenpaare

(B) TYP: Nukleinsäure

(C) STRANG: einzel

(D) TOPOLOGIE: linear

## (ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung hergestellte partielle cDNA

## (iii) HYPOTHETISCH: NEIN

## (iii) ANTI-SENSE: NEIN

## (vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(C) ORGAN:

## (vii) SONSTIGE HERKUNFT:

## (A) BIBLIOTHEK: cDNA library

## 5 (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 41

ctcgagccga attcggctcg agaggagccc agccctggga ttttcaggtg ttttcatttg 60  
 gtgatcagga ctgaacagag agaactcacc atggagtgtg ggctgagctg gctttttctt120  
 10 gtggtatttt taaaagggtg ccagtgtgag gtgcagctgt tggagtctgg gggaggcttg180  
 gtacagcctg gggggtcctt gagactctcc tgtgcagcct ctggattcac ctttagcagc240  
 tatgccatga gctgggtccg ccaggctcca ggaaggggc tggagtgggt ctcaggtatt300  
 agtggtagtg gtgtgatagt acacactacg caga 334

## 15 (2) INFORMATION ÜBER SEQ ID NO: 42:

## (i) SEQUENZ CHARAKTERISTIK:

- 20 (A) LÄNGE: 845 Basenpaare  
 (B) TYP: Nukleinsäure  
 (C) STRANG: einzel  
 (D) TOPOLOGIE: linear

25 (ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung  
 hergestellte partielle cDNA

30 (iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

## 35 (vi) HERKUNFT:

- (A) ORGANISMUS: MENSCH  
 (C) ORGAN:

## 40 (vii) SONSTIGE HERKUNFT:

- (A) BIBLIOTHEK: cDNA library

## 45 (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 42

gcgttccctc cgcgagcta cttctttctt tccttttttt tttttttctg gctaacagaa 60  
 ttttattggt aaatcacaga aacttttagt caaaacaaaa atcacgaagt ccatttaata120  
 gcaacttcat gtcctgctgg ctttgcttgc tgtctcctgg caaccagaag tggacagaag180  
 50 cgtgggtgcc caagtgggcc acagacagct tccaaccccc acaccagc atccaatcca240  
 caccagcag acccttcggc atgccgccct ctaccaggaa gccagaggcc taggagctcg300  
 ccatccatat ttatttgaaa aggtcaaaag gagcatctat gagacaaggg aggggtgcag360  
 gctgaagcag cgcctcaaca gccagggaca ttaggcaac acgagcagc acagcgggc420  
 caccactgtc cacacgtca cacaagccag gcccgaggc ccttcggaga gctagcaggt480  
 55 tacattcagg cagatggccc tcttcccacc caaacccaca gaaccccaaa caaggcatca540  
 ccaggaaaga cagggaag ccaaatcaca gttgaaccag ggacagagaa cccttggccc600  
 cactgatgtc ccaagccacc agcagctgtc tccaaaatcc ctatgctatt acagtggga660  
 ttacatcatt taaaagcct gattattccc aggttctaa tctttcatat aaaactgcct720  
 60 ttgttttgc cttttgttca actcagaggc ccagcaaagc gggcagggtc cctgatcagg780  
 gcaggagccc acctcagaag cccatgccgc accagtgtccc aagcacatgt cagtgtcag840  
 aacaa 845

## 65 (2) INFORMATION ÜBER SEQ ID NO: 43:

## (i) SEQUENZ CHARAKTERISTIK:

(A) LÄNGE: 2233 Basenpaare  
 (B) TYP: Nukleinsäure  
 (C) STRANG: einzel  
 (D) TOPOLOGIE: linear

5

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung  
 hergestellte partielle cDNA

10

(iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

15

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH  
 (C) ORGAN:

20

(vii) SONSTIGE HERKUNFT:

(A) BIBLIOTHEK: cDNA library

25

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 43

gaattcagaa	gttaatgatg	ttgggtaaga	gaacaatggt	aagagagcaa	tctaagaata	60
tatcacctac	tttaatttta	tatgagagta	catggaggta	gctgtgatgt	ggaaatgtag	120
cactgctcct	acccacgcag	atttattcca	gtgaaacaac	aactggaact	tcaagtaact	180
cctcccagag	tacttccaac	tctgggttgg	ccccaaatcc	aactaatgcc	accaccaagg	240
cggctggttg	tgccctgcag	tcaacagcca	gtctcttcgt	ggtctcactc	tctcttctgc	300
atctctactc	ttaagagact	caggccaaga	aacgtcttct	aaatttcccc	atcttctaaa	360
cccaatccaa	atggcgtctg	gaagtccaat	gtggcaagga	aaaacagggtc	ttcatcgaat	420
ctactaattc	cacacctttt	attgacacag	aaaatggtga	gaatcccaaa	tttgattgat	480
ttgaagaaca	tgtgagaggt	ttgactagat	gatggatgcc	aatattaaat	ctgctggagt	540
ttcatgtaca	agatgaagga	gaggcaacat	ccaaaatagt	taagacatga	tttccttgaa	600
tgtggcttga	gaaatatgga	cacttaatac	taccttgaaa	ataagaatag	aaataaagga	660
tgggattgtg	gaatggagat	tcagttttca	tttggttcat	taattctata	aggccataaa	720
acaggtaata	taaaaagctt	ccatgattct	atttatatgt	acatgagaag	gaacttccag	780
gtgttactgt	aattcctcaa	cgtattgttt	cgacagcact	aatttaatgc	cgatatactc	840
tagatgaagt	tttacattgt	tgagctattg	ctgttctctt	gggaactgaa	ctcactttcc	900
tcctgaggct	ttggatttga	cattgcattt	gaccttttat	gtagtaattg	acatgtgcca	960
gggcaatgat	gaatgagaat	ctacccccag	atccaagcat	cctgagcaac	tcttgattat	1020
ccatattgag	tcaaatggta	ggcatttctc	atcacctggt	tccattcaac	aagagcacta	1080
cattcattta	gctaaacgga	ttccaaagag	tagaattgca	ttgaccgcga	ctaatttcaa	1140
aatgcttttt	attattatta	tttttttagac	agtctcactt	tgtcgcgccag	gccggagtgc	1200
agtgggtcga	tctcagatca	gtgtaccatt	tgcctcccg	gtcgaagcga	ttctctgcc	1260
tcagcctccc	aagtagctgg	gattacaggc	acctgccacc	atgcccggt	aatttttgta	1320
attttagtag	agacagggtt	tcaccatggt	gcccgagctg	gtttcgaaact	cctgacctca	1380
ggtgatccac	ccgcctcggc	ctcccaaagt	gctgggatta	caggcttgag	cccccgcgcc	1440
cagccatcaa	aatgcttttt	atttctgcat	atgttgaata	ctttttacaa	tttaaaaaaa	1500
tgatctgttt	tgaaggcaaa	attgcaaata	ttgaaattaa	gaaggcaaaa	atgtaaagga	1560
gtcaaaaacta	taaatcaagt	atttgggaag	tgaagactgg	aagctaattt	gcattaaatt	1620
cacaaacttt	tatactcttt	ctgtatatac	attttttttc	tttaaaaaac	aactatggat	1680
cagaatagcc	acatttgtaa	cactttttgt	tatcagtcaa	tatttttaga	tagttagaac	1740
ctggtcctaa	gcctaaaagt	gggcttgatt	ctgcagtaaa	tctttttaca	ctgcctcgac	1800
acacataaac	cttttttaaa	atagacactc	cccgaggtct	tttgttcgca	tggtcacaca	1860
ctgatgctta	gatgttccag	taatctaata	tggccacagt	agtcttgatg	accaaagtcc	1920
ttttttttcca	tcttttagaaa	actacatggg	aacaaacaga	tcgaacagtt	ttgaagctac	1980
tgtgtgtgtg	aatgaacact	cttgctttat	tccagaatgc	tgtacatcta	ttttggattg	2040
tatatattgt	ttgtgtattt	acgctttgat	tcatagtaac	ttcttatgga	attgatttgc	2100
attgaacaca	aactgtaaat	aaaaagaaat	ggctgaaaga	gcaaaaaaaa	aggagggcag	2160
gagagaggaa	aagggggagga	agaggagggg	ggaaagagaa	gggagagaga	aggaggggga	2220
aggaggtggg	ggg					2233

65



(2) INFORMATION ÜBER SEQ ID NO: 44:

- 5 (i) SEQUENZ CHARAKTERISTIK:  
 (A) LÄNGE: 243 Basenpaare  
 (B) TYP: Nukleinsäure  
 (C) STRANG: einzel  
 10 (D) TOPOLOGIE: linear
- (ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung  
 hergestellte partielle cDNA  
 15
- (iii) HYPOTHETISCH: NEIN
- (iii) ANTI-SENSE: NEIN  
 20
- (vi) HERKUNFT:  
 (A) ORGANISMUS: MENSCH  
 (C) ORGAN:  
 25
- (vii) SONSTIGE HERKUNFT:  
 (A) BIBLIOTHEK: cDNA library  
 30
- (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 44

35 ggagcccagc actagaagtc ggcgggtgttt ccattcgggtg atcagcactg aacacagagg 60  
 actcaccatg gagtttgggc tgagctgggt ttctctcggt gctcttttaa gaggtgtcca120  
 gtgtcagggt cactgggtgga gcggggagcg ggtcagcagg agtcctgaat cctgtgacgc180  
 tgatcagtc tatatcagat ggcgcagctc agcagggtga tgggggtatga atgataacat240  
 40  
 243

(2) INFORMATION ÜBER SEQ ID NO: 45:

- 45 (i) SEQUENZ CHARAKTERISTIK:  
 (A) LÄNGE: 817 Basenpaare  
 (B) TYP: Nukleinsäure  
 (C) STRANG: einzel  
 50 (D) TOPOLOGIE: linear
- (ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung  
 hergestellte partielle cDNA  
 55
- (iii) HYPOTHETISCH: NEIN
- (iii) ANTI-SENSE: NEIN  
 60
- (vi) HERKUNFT:  
 (A) ORGANISMUS: MENSCH  
 (C) ORGAN:  
 65
- (vii) SONSTIGE HERKUNFT:  
 (A) BIBLIOTHEK: cDNA library

## (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 45

```

gttttttttt tttttttttt aagagagcag attctcttta ttgagatacg ggacacagcg 60
aagggtggag agacggaaca gcccccagc ctcagccctc tccacggggg ccggatgccal20
gagatgggag aagggtattca gtctctcgcc cgggaaaccc agtcccacag agggcgccgg180
caaggggtggg acgcgacctg ggtgacacgg tgcaggaggt ctttaaatag aggaggggct240
ggagcgggga aacgcgcggg ggccctagcg caccatgtat tccttgcgct tattgagccg300
aacttggcag aaagagaagc ctccgaggag gaggtaaagg cctgcagcga tgaacagtt360
gtagctgact tgctcgtaaa ggttgatat gttctggggg ccattctcaa aatctttctc420
cgtgaaggga acgtcctcaa tcaacacagc ggaatggaca ttgaaaaata ttccgagcat480
tatcaacatg atcactcccc aggcgctgag gacgatgccg caggcgccca gcttcggccc540
acagcacagg agcgacgcca taaagaaggg agtcggggat cgccgagggt caagcgggct600
cggaaagcgg tgggagaaaag cccaggatgc cctcgcaggg gggcagaggg ggcgtggccc660
cggcctcaac catcccatcc gggggcgcca ggcggaagg gctgggctcc tctcaggact720
ttcgcgggag acggcgccgt ctgaaaccaa aactgctcct ggggaaacct tccttgacct780
ctgtagctag ggcgtgagta ttggaagagc gagggcc 817

```

## (2) INFORMATION ÜBER SEQ ID NO: 46:

## (i) SEQUENZ CHARAKTERISTIK:

(A) LÄNGE: 1644 Basenpaare

(B) TYP: Nukleinsäure

(C) STRANG: einzel

(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung  
hergestellte partielle cDNA

(iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(C) ORGAN:

(vii) SONSTIGE HERKUNFT:

(A) BIBLIOTHEK: cDNA library

## (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 46

```

gttcgggtc acatgggaaa tactttctga ggtcctgga cctcctgtgc aagaacatga 60
aacacctgtg gttcttcctc ctgctgggtg cagctcccag atgggtcctg tcccagggtg 120
agctgcagga gtcgggecca ggactgggtg agccttcaca gacctgtcc ctcacctgca 180
ctgtctctgg tggctccatc agcagtggtg gttactactg gagctggatc cgccagcacc 240
cagggaaagg cctggagtggt attgggtaca tctattacag tgggagcacc tactacaacc 300
cgtcctctca gagtcgagtt accatatcag tagacacgtc taagaaccag ttctccctga 360
agctgagctc ttgtactgcc gcggacacgg ccgtgtatta ctgtgcgaga gagcatctct 420
cctacgggtg ctcgagatac tactactacg gtatggacgt ctggggccaa gggaccgggt 480
caccgtctcc tcagcatccc cgaccagccc caaggtcttc ccgctgagcc tctgcagcac 540
ccagccagat gggaaacgtg tcatcgctg cctggtccag ggcttcttcc ccaggagacc 600
actcagtgtg acctggagcg aaagggacag ggcgtgaccg ccagaaactt cccaccacgc 660
caggatgcct ccggggacct gtacaccacg agcagccagc tgacctgcc ggccacacag 720
tgcttagccg gcaagtcctg gacatgccac gtgaagcact acacgaatcc cagccaggat 780
gtgactgtgc cctgccagt tccctcaact ccacctaccc catctccctc aactccacct 840
accccatctc cctcatgctg ccacccccga ctgtcactgc accgaccggc cctcgaggac 900

```

ctgctcttag gttcagaagc gaacctcagc tgcacactga ccggcctgag agatgcctca 960  
 ggtgtcacct tcacctggac gccctcaagt gggaaagacg ctgttcaagg accacctgag1020  
 cgtgacctct gtggctgcta cagcgtgtcc agtgtcctgc cgggctgtgc cgagccatgg1080  
 5 aacctaggga agaccttcac ttgactgtct gcctaccccg agtccaagac cccgctaacc1140  
 gccaccctct caaaatccgg aaacacattc cggcccgagg tccacctgct gccgccgccc1200  
 tcggaggagc tggccctgaa cgagctgggtg acgctgacgt gcctggcacg cggcttcagc1260  
 cccaaggacg tgctggttcg ctggctgcag gggtcacagg agctgccccg cgagaagtac1320  
 ctgacttggg catcccggca ggagcccagc cagggcacca ccaccttcgc tgtgaccagc1380  
 10 atactgcgcg tggcagccga ggactggaag aagggggaca cttctcctg catggtgggc1440  
 cagaggccc tggcgtggc cttcacacag aagaccatcg accgcttggc gggtaaacccl1500  
 acccatgtca atgtgtctgt tgtcatggcg gaggtggacg gcacctgcta ctgagccgccc1560  
 cgccctgtcc caccctgaa taaactccat gctccccca gcaaaaaaaa aaaaaaaaaa1620  
 15 aaaaaaaaaa aaaaaaaaaa aaaa 1644

## (2) INFORMATION ÜBER SEQ ID NO: 47:

20

## (i) SEQUENZ CHARAKTERISTIK:

(A) LÄNGE: 1133 Basenpaare

(B) TYP: Nukleinsäure

25

(C) STRANG: einzel

(D) TOPOLOGIE: linear

30

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung  
hergestellte partielle cDNA

(iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(C) ORGAN:

(vii) SONSTIGE HERKUNFT:

(A) BIBLIOTHEK: cDNA library

45

## (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 47

atttatctgg gacagacatc ttcagaatga cacatgccaa acagtgggtc ttattaaatc 60  
 50 aaaggttcag atattatcag attcagaaat agtgatgctt tgtgtatcta tttctctctc 120  
 tttaaacaga aaaagacaaa tgaatgggga aagacaatca ttgaatacaa aacaaataag 180  
 ccatcacgcc tgccttctct tgatattgca cctttggaca tcggtgggtc tgaccaggaa 240  
 ttctttgtgg acattggccc agtctgtttc aaataaatga actcaatcta aattaaaaaa 300  
 55 gaaagaaatt tgaaaaaact ttctctttgc catttcttct tcttcttttt taactgaaag 360  
 ctgaatcctt ccatttcttc tgcacatcta cttgtttaa ttgtgggcaa aagagaaaaa 420  
 gaaggattga tcagagcatt gtgcaataca gtttcattaa ctcttcccc cgctccccc 480  
 aaaatttgaa ttttttttcc aacactctta cacctgttat ggaaaatgtc aacctttgta 540  
 agaaaaccaa aataaaaatt gaaaaataaa aaccataaac atttgacca cttgtggctt 600  
 60 ttgaatatct tccacagagg gaagtttaaa acccaactt ccaaagggtt aaactacctc 660  
 aaaacacttt cccatgagtg tgatccacat tgtaggtgc tgacctagac agagatgaac 720  
 tgaggtcctt gttttgtttt gttcataata caaagggtct aattaatagt atttcagata 780  
 cttgaagaat gttgatgggt ctagaagaa ttgagaagaa atactcctgt attgagttgt 840  
 atcgtgtggt gtatttttta aaaaatttga tttagcattc atattttcca tcttattccc 900  
 65 aattaaaagt atgcagatta tttgcccata tcttcttcag attcagcatt tgttctttgc 960  
 cagtctcatt ttcattctct tccatgggtc cacagaagct ttgtttcttg ggcaagcaga1020  
 aaaattaaat tgtacctatt ttgtatatgt gagatgttta aataaattgt gaaaaaatg1080  
 aaataaagca tgtttggttt tccaaaagaa aaaaaaaaaa aaaaaagtcg acc 1133

## (2) INFORMATION ÜBER SEQ ID NO: 48:

## (i) SEQUENZ CHARAKTERISTIK:

- (A) LÄNGE: 969 Basenpaare  
 (B) TYP: Nukleinsäure  
 (C) STRANG: einzel  
 (D) TOPOLOGIE: linear

## (ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung hergestellte partielle cDNA

## (iii) HYPOTHETISCH: NEIN

## (iii) ANTI-SENSE: NEIN

## (vi) HERKUNFT:

- (A) ORGANISMUS: MENSCH  
 (C) ORGAN:

## (vii) SONSTIGE HERKUNFT:

- (A) BIBLIOTHEK: cDNA library

## (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 48

```

gaggaggagg gtgtatctcc tttcgtcgga ccgccccttg gcttctgcac tgatggtggg 60
tggatgagta atgcatccag gaagcctgga ggcctgtggt ttccgcaccc gctgccaccc120
ccgcccctag cgtggacatt tatcctctag cgctcaggcc ctgccgccat cgccgcagat180
ccagcgccca gagagacacc agagaaccga ccattggcccc ctttgagccc ctggcttctg240
gcatcctggt gttgctgtgg ctgatagccc ccagcagggc ctgcacctgt gtcccacccc300
accacagac ggccttctgc aattccgacc tcgtcatcag ggccaagtgc gtggggacac360
cagaagtcaa ccagaccacc ttataccagc gttatgagat caagatgacc aagatgtata420
aagggttcca agccttaggg gatgccgctg acatccggtt cgtctacacc ccgccatgg480
agagtgtctg cggatacttc cacaggtccc acaaccgcag cgaggagttt ctcatgtctg540
gaaaactgca ggtggactc ttgcatatca ctacctgcag tttcgtggct ccctggaaca600
gcctgagctt agctcagcgc cggggcttca ccaagacctc cactgttggc tgtgaggaaat660
gcacagtgtt tccctgttta tccatccctt gcaaaactgca gagtggcact cattgcttgt720
ggacggacca gctcctccaa ggctctgaaa agggcttcca gtcccgtcac cttgcctgcc780
tgccctcgga gccagggtg tgccacctggc agtccctgcg gtcccagata gcctgaatcc840
tgcccgaggt ggaagctgaa gcctgcaag tgtccacctt gttccactc ccattcttct900
tccggacaat gaaataaaga gttaccaccc agcaaaaaaa aaaaaaaaaa acaagtcgtc960
gcgtgctgt 969

```

## (2) INFORMATION ÜBER SEQ ID NO: 49:

## (i) SEQUENZ CHARAKTERISTIK:

- (A) LÄNGE: 617 Basenpaare  
 (B) TYP: Nukleinsäure  
 (C) STRANG: einzel  
 (D) TOPOLOGIE: linear

## (ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung hergestellte partielle cDNA

(iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(C) ORGAN:

(vii) SONSTIGE HERKUNFT:

(A) BIBLIOTHEK: cDNA library

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 49

```

cctacaccta cctccctttt gggtttctat tccgaccgcg atgatttgct ttggaaggct 60
taacccccctt cttccccaaa cttgcccccg gagaaccccc agccttacga ccctcctcct120
gaagatgcaa aaccagcttg cggcgccgcg tctcttccag gacatcaaga agccagctga180
agatgagtgg ggtaaaaccc cagacgccat gaaagctgcc atggccctgg agaaaaagct240
gaaccagggc cttttggatc ttcattgccct gggttctgcc cgcacggacc cccatctctg300
tgacttcctg gagactcact tcctagatga ggaagtgaag cttatcaaga agatgggtga360
ccacctgacc aacctccaca ggctgggtgg cccggaggct gggctgggcg agtatctctt420
cgaaaggctc actctcaagc acgactaaga gccttctgag cccagcgact tctgaagggc480
cccttgcaaa gtaatagggc ttctgcctaa gcctctccct ccagccaata ggcagctttc540
ttaactatcc taacaagcct tggaccaaag ggaataaag ctttttgatg caaaaaaaga600
ggagggggga aaaaagc                                     617

```

(2) INFORMATION ÜBER SEQ ID NO: 50:

(i) SEQUENZ CHARAKTERISTIK:

(A) LÄNGE: 704 Basenpaare

(B) TYP: Nukleinsäure

(C) STRANG: einzel

(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung hergestellte partielle cDNA

(iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(C) ORGAN:

(vii) SONSTIGE HERKUNFT:

(A) BIBLIOTHEK: cDNA library

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 50

```

ggggagactc gtcaccaggc gtgcagtggg cactgctggg ctcccccatc ccgtcctaac 60
ccggaacagc cccgggcagg aggcgtggaa agtcgagggg gtaaacccgc aatgtgcgtt120
gtgtaagcca cggcgacagg tggggcgagg gcgggacttg ggcgggcggg gtgggcttgg180
ccgagctggc ctccggggca ccgaccgcta taaggccagt cggactgcga cacagcccat240
ccctcgacc gctcgcgtcg catttgcccg cctccctacc gctccaagcc cagccctcag300
ccatggcatg cccctggat caggccattg gcctcctcgt ggccatcttc cacaagtact360

```

ccggcagggg ggggtgacaag cacaccctga gcaagaagga gctgaacgag ctgatccaga420  
 aggagctcac cattggctcg aagctgcagg atgctgaaat tgcaaggctg atggaagact480  
 tggaccggaa caaggaccag gaggtgaact tccaggagta tgtcaccttc ctggggggcct540  
 tggctttgat ctacaatgaa gccctcaagg gctgaaaata aatagggaag atggagacac600  
 cctctggggg tcctctctga gtcaaatcca gtgggtgggt attgtacaat aaattttttt660  
 tgggtcaaatt taataaaaaa aaaaaaagag aaaaaagggt gagg 704

(2) INFORMATION ÜBER SEQ ID NO: 51:

- (A) LÄNGE: 95 Aminosäuren
- (B) TYP: Protein
- (C) STRANG: einzel
- (D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 51:

PCSSQFHPVE NRSQEPLAGD SMSPTLPVQ NMNNAMFLQK TLSLSFIGGN HQTAECTRL60  
 SRTTDLSPSH SPYHHKSHNK KEKRYFGFKK SKKIM 95

(2) INFORMATION ÜBER SEQ ID NO: 52:

- (A) LÄNGE: 76 Aminosäuren
- (B) TYP: Protein
- (C) STRANG: einzel
- (D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 52:

LPRDTWFKLK CLTDHSRHLV HSYVNVSHLT WWHCLQTEHR LPLAWFENRN RAMPTDPSYV60  
 WASKWNCTFI QIFTCL 76

(2) INFORMATION ÜBER SEQ ID NO: 53:

- (A) LÄNGE: 90 Aminosäuren
- (B) TYP: Protein
- (C) STRANG: einzel
- (D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

5

(vi) HERKUNFT:  
(A) ORGANISMUS: MENSCH

:

10

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 53:  
RVNNCQEQLV VLKYNPQPRI PPVLQMDQLK QANTEDTKNE VRFIETRVTP LDELNTKMTL60  
TLSTRYSSET CLQNEIPEEF CSYPEIRGSN 90

15

(2) INFORMATION ÜBER SEQ ID NO: 54:

(A) LÄNGE: 117 Aminosäuren  
(B) TYP: Protein  
(C) STRANG: einzel  
(D) TOPOLOGIE: linear

20

25

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

30

(vi) HERKUNFT:  
(A) ORGANISMUS: MENSCH

:

35

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 54:  
PLQPPRAMAP RGCIVAVFAI FCISRLLC SH GAPVAPMPY LMLCQPHKRC GDKFYDPLQH 60  
CCYDDAVVPL ARTQTCGNCT FRVCFEQCCP WTFMVKLINQ NCDSARTSDD RLCRSVS 117

40

(2) INFORMATION ÜBER SEQ ID NO: 55:

(A) LÄNGE: 103 Aminosäuren  
(B) TYP: Protein  
(C) STRANG: einzel  
(D) TOPOLOGIE: linear

45

50

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

55

(vi) HERKUNFT:  
(A) ORGANISMUS: MENSCH

:

60

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 55:  
RVQPALGWC RVDVGHGRHQ EGSESLGPHQ HTHMLSRIL EGDWASSGQ RQGGPQTGHR 60  
MKWAVECVFL WPPNSHSASQ ISGNTSLFLQ AHPGRRIQES SFP 103

65

(2) INFORMATION ÜBER SEQ ID NO: 56:

(A) LÄNGE: 81 Aminosäuren  
(B) TYP: Protein  
(C) STRANG: einzel  
(D) TOPOLOGIE: linear

5

(ii) MOLEKÜLTYP: ORF

10

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:

15

(A) ORGANISMUS: MENSCH

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 56:

20

RCSFHTSGSW PRARRHHHSN SAAGGRRTCP HISCVAGTAS GKESWGPLGL RVSRGAWRCR60  
KWQRQLRCSL GEPWLWVVAV E 81

(2) INFORMATION ÜBER SEQ ID NO: 57:

25

(A) LÄNGE: 125 Aminosäuren  
(B) TYP: Protein  
(C) STRANG: einzel  
(D) TOPOLOGIE: linear

30

(ii) MOLEKÜLTYP: ORF

35

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:

40

(A) ORGANISMUS: MENSCH

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 57:

45

RAARADSARA FPLPACKVVV PQGPPPGHVG AAGQAFPSFE RGFRCSRAS GLRSSLPSPR 60  
SVVASPPPTH QSRCILGRAL GAMAPRGRKR KAEAAVVAVA EKREKLANGG EGMEEATVVI120  
EHCTS 125

(2) INFORMATION ÜBER SEQ ID NO: 58:

50

(A) LÄNGE: 119 Aminosäuren  
(B) TYP: Protein  
(C) STRANG: einzel  
(D) TOPOLOGIE: linear

55

(ii) MOLEKÜLTYP: ORF

60

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:

65

(A) ORGANISMUS: MENSCH



(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 58:

5 QRSPPFPFPPR RSPASLASRL RRPPRPQPYA SSRGEPWRLE PGRECSGTGG WGAETRPLSG 60  
NWATKSAARK LCSYSGNLSQ RKGKLGPHQHP RGLADLGAQ PLCKQGAGRL EPNRLERLE 119

(2) INFORMATION ÜBER SEQ ID NO: 59:

- 10 (A) LÄNGE: 128 Aminosäuren  
(B) TYP: Protein  
(C) STRANG: einzel  
15 (D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

20 (iii) HYPOTHETISCH: ja

(vi) HERKUNFT:

25 (A) ORGANISMUS: MENSCH

:

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 59:

30 TRQLVQCSIT TVASSIPSP FASFSRFSAT ATTAASALRF LPRGAMAPRA RPRMQRDWWV 60  
GGGDATERK LGNEERSPEA LLLQRKPLSK EGKAWPAAPT WPGGGPWGTT TLQAGSGKAR120  
AESARAAR 128

35 (2) INFORMATION ÜBER SEQ ID NO: 60:

- (A) LÄNGE: 127 Aminosäuren  
(B) TYP: Protein  
40 (C) STRANG: einzel  
(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

45 (iii) HYPOTHETISCH: ja

50 (vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

:

55 (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 60:

VTVMQNFEL SFKYVLYSSY SWLKL DHTIA DCMVFTWTPC RMLDYLYSSY ANMLWAGEMK 60  
SSSHQDLLFK WLDNWATKEL ELHLLGFELF WNTLLHFGKS KSSASGALSI ENLPSFALKD120  
VLFFIYT 127

60

(2) INFORMATION ÜBER SEQ ID NO: 61:

- (A) LÄNGE: 111 Aminosäuren  
65 (B) TYP: Protein  
(C) STRANG: einzel  
(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

5

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

10

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 61:

SIGPICSQGL GPGGIPSPIT LIKNGCNCKN PCLIIYLQCS HLQMYLLMLS CQVPMQRWRG 60  
LPLCGWGLWV VVKDRYQKNA FKCTNLLINI RCLLKKKKKK KKRVGVGCI G 111

15

(2) INFORMATION ÜBER SEQ ID NO: 62:

(A) LÄNGE: 68 Aminosäuren

20

(B) TYP: Protein

(C) STRANG: einzel

(D) TOPOLOGIE: linear

25

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

30

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

35

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 62:

YRSFTTTHKP HPHKGSPRHL CIGTWQLSIR RYICKWEHSC KYIRQGFLQL QPFLIKVIGE60  
GIPPGPRP 68

40

(2) INFORMATION ÜBER SEQ ID NO: 63:

(A) LÄNGE: 195 Aminosäuren

45

(B) TYP: Protein

(C) STRANG: einzel

(D) TOPOLOGIE: linear

50

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

55

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

60

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 63:

LVQPGGSCSG GRLLGVEFPS APRVRPFERS APAPATSLLG AMTTTTTFKG VDPNSRNSSR 60  
VLRPPGGGSN FSLGFDEPTE QPVRKNKMAS NIFGTPEENQ ASWAKSAGAK SSGGREDLES120  
SGLQRRNSSE ASSGDFDLK GEGDIHENVD TDLPGSLGQS EEKPVPAAPV PSPVAPAPVP180

65

SRRNPPGGKS SLVLG

## (2) INFORMATION ÜBER SEQ ID NO: 64:

5

- (A) LÄNGE: 164 Aminosäuren
- (B) TYP: Protein
- (C) STRANG: einzel
- (D) TOPOLOGIE: linear

10

(ii) MOLEKÜLTYP: ORF

15

(iii) HYPOTHETISCH: ja

20

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

:

25

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 64:

VSQSFPSNLL LENTHAMHR PKSQQQRETC SSKEKKKRQQ YIKCFFLMKQ IQEMYSQAQV 60  
 VQFTSMEETD RTTAFRTVRA NRRGWTCRQ GDFFWMALGP GPPGWAQAQQ ARASLHSAPG120  
 CLASLCPHFH EYHLLPSDLR SLRSLLRSS FSAVQMTPSL PCHH 164

30

## (2) INFORMATION ÜBER SEQ ID NO: 65:

35

- (A) LÄNGE: 106 Aminosäuren
- (B) TYP: Protein
- (C) STRANG: einzel
- (D) TOPOLOGIE: linear

40

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

45

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

:

50

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 65:

FQAVSLYIQA FLCVRAKALL ISQPVLLLSG YFLRLKNKRQ FLCFAGGKAG GAGLFIVHMS 60  
 QEEALSKGHW QVRATPRRLC GETPCGLGPG RNGACGLFMV CPVEAW 106

55

## (2) INFORMATION ÜBER SEQ ID NO: 66:

60

- (A) LÄNGE: 349 Aminosäuren
- (B) TYP: Protein
- (C) STRANG: einzel
- (D) TOPOLOGIE: linear

65

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:  
(A) ORGANISMUS: MENSCH

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 66:

AALRSDAGMK RALGRRKGWV LRLRKILFCV LGLYIAIPFL IKLCPGIQAK LIFLNFVRVP 60  
YFIDLKKPQD QGLNHTCNYY LQPEEDVTIG VWHTVPAVWW KNAQGKQMW YEDALASSHP120  
IILYLHGNAG TRGGDHRVEL YKVLSSLGYH VVTFDYRGWG DSVGTPSERG MTYDALHVFD180  
WIKARSGDNP VYIWGHSLGT GVATNLVRRL CERETPPDAL ILESFFTNR EEAQSHPPSV240  
IYRYFPGFDW FFLDPITSSG IKFANDENVK HISCPILLIH AEDDPVVPFQ LGRKLYSIAA300  
PARSFRDFKV QFVPFHSDLG YRHKYIYKSP ELPRILREFL GKSEPEHQH 349

(2) INFORMATION ÜBER SEQ ID NO: 67:

(A) LÄNGE: 191 Aminosäuren  
(B) TYP: Protein  
(C) STRANG: einzel  
(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:  
(A) ORGANISMUS: MENSCH

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 67:

SGLSRLGPGR NQHAGQDVLC EVAAALHQVL KELLGQGIDY EKILKLTADA KFESGDVKAT 60  
VAVLSFILSS AAKHSVDGES LSSELQQLGL PKEHAASLCR CYEEKQSPLQ KHLRVCSLRM120  
NRLAGVGWRV DYTLLSSLLQ SVEEPMVHLR LEVAAAPGTP AQPVAMSLSA DKFQVLLAEL180  
KQAQTLMSSL G 191

(2) INFORMATION ÜBER SEQ ID NO: 68:

(A) LÄNGE: 164 Aminosäuren  
(B) TYP: Protein  
(C) STRANG: einzel  
(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:  
(A) ORGANISMUS: MENSCH

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 68:

FFFFFFFFF FFFFSLLYFC LFFLLMKTAN NCLSREGKVM LGKVLRSPEP SSQERSEAAG 60  
DLGGQSPGQG LSILEPGLPP EEQFRGRDSI RAGRLHTGLE HPSPQPRELI RVWACFSSAR120

RTWNLSAERD MATGWAGVPG AAATSSRRCT MGSSTDCSRL ELRV

164

## (2) INFORMATION ÜBER SEQ ID NO: 69:

- (A) LÄNGE: 155 Aminosäuren  
(B) TYP: Protein  
(C) STRANG: einzel  
(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

## (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 69:

NQGCLPKSSS EGVTPYQGG STQAWNTLLL SPGSSSGSGP ASVLPGGPGT CLLRGTWQQA 60  
GLGSLGQLPP PAAGAPWALP RIAAGWSSGC SPPASPHLPT YSCVGCRPAS ASARGFASPH120  
NSGTGWPRAL WAAPAAVHW TRIRHRHCAW PHWRG 155

## (2) INFORMATION ÜBER SEQ ID NO: 70:

- (A) LÄNGE: 35 Aminosäuren  
(B) TYP: Protein  
(C) STRANG: einzel  
(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

## (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 70:

RRAAVTWVWL GVLCFESAVF TPTEVVRTCR LLRFS

35

## (2) INFORMATION ÜBER SEQ ID NO: 71:

- (A) LÄNGE: 32 Aminosäuren  
(B) TYP: Protein  
(C) STRANG: einzel  
(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:  
(A) ORGANISMUS: MENSCH

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 71:  
KRLTQNTTPP TRPKSQLHVF KTSFKVSYFS TS

32

(2) INFORMATION ÜBER SEQ ID NO: 72:

(A) LÄNGE: 37 Aminosäuren  
(B) TYP: Protein  
(C) STRANG: einzel  
(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:  
(A) ORGANISMUS: MENSCH

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 72:  
ENRSNLHVLT TSVGKVTADS KHNTPNQTQV TAARLQN

37

(2) INFORMATION ÜBER SEQ ID NO: 73:

(A) LÄNGE: 121 Aminosäuren  
(B) TYP: Protein  
(C) STRANG: einzel  
(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:  
(A) ORGANISMUS: MENSCH

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 73:  
LVKGMTVLEA VLEIQAITGS RLLSMVPGPA RPPGSCWDPT QCTRTWLLSH TPRRRWISGL 60  
PRASCRLGEE PPPLPYCDQA YGEELSIRHR ETWAWLSRTD TAWPGAPGVK QARILGELLL120  
V 121

(2) INFORMATION ÜBER SEQ ID NO: 74:

(A) LÄNGE: 115 Aminosäuren  
(B) TYP: Protein  
(C) STRANG: einzel  
(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

5 (iii) HYPOTHETISCH: ja

(vi) HERKUNFT:

10 (A) ORGANISMUS: MENSCH

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 74:

15 QACPWASLAQ GQTRLRRKL DTPVHGGLGL EGWLSGLEVP GGLPAGTRPS AAGWAVPCCC 60  
CPQGLAVVAE DGTLSGWIRS PGSSSSRELR HKAGARLYTC RTQESLLQFL PEAPR 115

(2) INFORMATION ÜBER SEQ ID NO: 75:

20 (A) LÄNGE: 117 Aminosäuren

(B) TYP: Protein

(C) STRANG: einzel

25 (D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

30 (iii) HYPOTHETISCH: ja

(vi) HERKUNFT:

35 (A) ORGANISMUS: MENSCH

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 75:

40 RWHLIRLDQV TRQQQLSRAE AQGRGPAVHL QDPGEPVAVL ARSAEIASSV SLQQEQNQLW 60  
PRWVGSAFL AMAAATPRQE TAECLEGCNT RSNRQPPLFL MSDGQALQHL DRHGGWS 117

(2) INFORMATION ÜBER SEQ ID NO: 76:

45 (A) LÄNGE: 66 Aminosäuren

(B) TYP: Protein

(C) STRANG: einzel

50 (D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

55 (iii) HYPOTHETISCH: ja

(vi) HERKUNFT:

60 (A) ORGANISMUS: MENSCH

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 76:

65 PPQERRTIFV LYPRGSGREN MESGFYRLIG PIHKGHDWEK VWEQKENWDF RVQYAHPKLL60  
VAWGMS 66

(2) INFORMATION ÜBER SEQ ID NO: 77:

- (A) LÄNGE: 81 Aminosäuren
- (B) TYP: Protein
- (C) STRANG: einzel
- (D) TOPOLOGIE: linear

5

(ii) MOLEKÜLTYP: ORF

10

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:

15

(A) ORGANISMUS: MENSCH

:

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 77:

20

ALSTRAMTGK RYGSRRRIGI LGCSTLTNLF WWPGACLEAQ TVKQALLACL LVTTSAPAVL60  
RLHPAPGTPP APEPPLSPCD G 81

(2) INFORMATION ÜBER SEQ ID NO: 78:

25

- (A) LÄNGE: 104 Aminosäuren
- (B) TYP: Protein
- (C) STRANG: einzel
- (D) TOPOLOGIE: linear

30

(ii) MOLEKÜLTYP: ORF

35

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:

40

(A) ORGANISMUS: MENSCH

:

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 78:

45

TLTIIHIWTR DTVHEVHPSQ GDSGGSGAGG VPGAGWSLKT AGAEVVTISKQ ASRACLTVWA 60  
SRHAPGHQKF RVSVLHPKIP ILLLLPYLFP VMALVDRAVQ SIES 104

(2) INFORMATION ÜBER SEQ ID NO: 79:

50

- (A) LÄNGE: 104 Aminosäuren
- (B) TYP: Protein
- (C) STRANG: einzel
- (D) TOPOLOGIE: linear

55

(ii) MOLEKÜLTYP: ORF

60

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:

65

(A) ORGANISMUS: MENSCH



(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 79:

PSSPRAVRHS GHQDSKMASV VPVKDKKLE VKLGELPSWI LMRDFSPSGI FGAFQRGYYR 60  
YNNKYINVKK GSISGITMVL ACYVLFSYSF SYKHLKHERL RKYH 104

(2) INFORMATION ÜBER SEQ ID NO: 80:

- (A) LÄNGE: 82 Aminosäuren
- (B) TYP: Protein
- (C) STRANG: einzel
- (D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 80:

RRGASRGLPW CWHATCSLAT PFPTSISSTS GSANTTEEDT LCTPPPHDLG PSPSVRNTIS60  
IVAESFHILI GINLQIKHDW YV 82

(2) INFORMATION ÜBER SEQ ID NO: 81:

- (A) LÄNGE: 115 Aminosäuren
- (B) TYP: Protein
- (C) STRANG: einzel
- (D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 81:

KDSATIEIVF LTEGLGPRSW GGGVQSVSSS VVFAEPLVLE MLVGKGVAKE HVACQHHGNP 60  
RDAPLLHIDV LVVVPVVTSL KRSENATRTE VPHQDPAWQL PQFDLQKFLV LHWYN 115

(2) INFORMATION ÜBER SEQ ID NO: 82:

- (A) LÄNGE: 187 Aminosäuren
- (B) TYP: Protein
- (C) STRANG: einzel
- (D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

- b) eine allelische Variation der unter a) genannten Nukleinsäure-Sequenzen  
oder
- c) eine Nukleinsäure-Sequenz, die komplementär zu den unter a) oder b) genannten Nukleinsäure-Sequenzen ist.
- 2. Eine Nukleinsäure-Sequenz gemäß einer der Sequenzen Seq. ID Nos. 1–50, oder eine komplementäre oder allelische Variante davon. 5
- 3. Nukleinsäure-Sequenz Seq. ID No. 1 bis Seq. ID No. 50, dadurch gekennzeichnet, daß sie in Blasen-tumorge-webe erhöht exprimiert sind.
- 4. BAC, PAC und Cosmid-Klone, enthaltend funktionelle Gene und ihre chromosomale Lokalisation, entsprechend den Sequenzen Seq. ID. No. 1 bis Seq. ID No. 50, zur Verwendung als Vehikel zum Gentransfer. 10
- 5. Eine Nukleinsäure-Sequenz gemäß den Ansprüchen 1 bis 4, dadurch gekennzeichnet, daß sie eine 90%ige Ho-mologie zu einer humanen Nukleinsäure-Sequenz aufweist.
- 6. Eine Nukleinsäure-Sequenz gemäß den Ansprüchen 1 bis 4, dadurch gekennzeichnet, daß sie eine 95%ige Ho-mologie zu einer humanen Nukleinsäure-Sequenz aufweist.
- 7. Eine Nukleinsäure-Sequenz, umfassend einen Teil der in den Ansprüchen 1 bis 6 genannten Nukleinsäure-Se- quenzen, in solch einer ausreichenden Größe, daß sie mit den Sequenzen gemäß den Ansprüchen 1 bis 6 hybridisie- ren. 15
- 8. Ein Nukleinsäure-Sequenz gemäß den Ansprüchen 1 bis 7, dadurch gekennzeichnet, daß die Größe des Frag- ments eine Länge von mindestens 50 bis 4500 bp aufweist.
- 9. Eine Nukleinsäure-Sequenz gemäß den Ansprüchen 1 bis 7, dadurch gekennzeichnet, daß die Größe des Frag- ments eine Länge von mindestens 50 bis 4000 bp aufweist. 20
- 10. Eine Nukleinsäure-Sequenz gemäß einem der Ansprüche 1 bis 9, die mindestens eine Teilsequenz eines biolo- gisch aktiven Polypeptids kodiert.
- 11. Eine Expressionskassette, umfassend ein Nukleinsäure-Fragment oder eine Sequenz gemäß einem der Ansprü- che 1 bis 9, zusammen mit mindestens einer Kontroll- oder regulatorischen Sequenz. 25
- 12. Eine Expressionskassette, umfassend ein Nukleinsäure-Fragment oder eine Sequenz gemäß Anspruch 11, wo- rin die Kontroll- oder regulatorische Sequenz ein geeigneter Promotor ist.
- 13. Eine Expressionskassette gemäß einem der Ansprüche 11 und 12, dadurch gekennzeichnet, daß die auf der Kas- sette befindlichen DNA-Sequenzen ein  
Fusionsprotein kodieren, das ein bekanntes Protein und ein biologisch aktives Polypeptid-Fragment umfaßt. 30
- 14. Verwendung der Nukleinsäure-Sequenzen gemäß den Ansprüchen 1 bis 10 zur Herstellung von Vollängen-Ge- nen.
- 15. Ein DNA-Fragment, umfassend ein Gen, das aus der Verwendung gemäß Anspruch 14 erhältlich ist.
- 16. Wirtszelle, enthaltend als heterologen Teil ihrer exprimierbaren genetischen Information ein Nukleinsäure- Fragment gemäß einem der Ansprüche 1 bis 10. 35
- 17. Wirtszelle gemäß Anspruch 16, dadurch gekennzeichnet, daß es ein prokaryontisches oder eukaryontische Zell- system ist.
- 18. Wirtszelle gemäß einem der Ansprüche 16 oder 17, dadurch gekennzeichnet, daß das prokaryontische Zellsy- stem E. coli und das eukaryontische Zellsystem ein tierisches, humanes oder Hefe-Zellsystem ist.
- 19. Ein Verfahren zur Herstellung eines Polypeptids oder eines Fragments, dadurch gekennzeichnet, daß die Wirts- zellen gemäß den Ansprüchen 16 bis 18 kultiviert werden. 40
- 20. Ein Antikörper, der gegen ein Polypeptid oder ein Fragment gerichtet ist, welches von den Nukleinsäuren der Sequenzen Seq. ID No. 1 bis Seq. ID No. 50 kodiert wird, das gemäß Anspruch 19 erhältlich ist.
- 21. Ein Antikörper gemäß Anspruch 20, dadurch gekennzeichnet, daß er monoklonal ist.
- 22. Ein Antikörper gemäß Anspruch 20 dadurch gekennzeichnet, daß er ein Phage-Display-Antikörper ist. 45
- 23. Polypeptid-Teilsequenzen, gemäß den Sequenzen Seq. ID Nos. ORF 51–106.
- 24. Polypeptid-Teilsequenzen gemäß Anspruch 22, mit mindestens 80%iger Homologie zu diesen Sequenzen.
- 25. Ein aus einem Phage-Display hervorgegangenen Polypeptid, welches an die Polypeptid-Teilsequenzen gemäß Anspruch 24 binden kann.
- 26. Polypeptid-Teilsequenzen gemäß Anspruch 22, mit mindestens 90%iger Homologie zu diesen Sequenzen. 50
- 27. Verwendung der Polypeptid-Teilsequenzen gemäß den Sequenzen Seq. ID No. 51–106, als Tools zum Auffin- den von Wirkstoffen gegen den Blasen-tumor.
- 28. Verwendung der Nukleinsäure-Sequenzen gemäß den Sequenzen Seq. ID No. 1 bis Seq. ID No. 50 zur Expres- sion von Polypeptiden, die als Tools zum Auffinden von Wirkstoffen gegen den Blasen-tumor verwendet werden können. 55
- 29. Verwendung der Nukleinsäure-Sequenzen Seq. ID No. 1 bis Seq. ID No. 50 in sense oder antisense Form.
- 30. Verwendung der Polypeptid-Teilsequenzen Seq. ID No. 51–106 als Arzneimittel in der Gentherapie zur Be- handlung des Blasen-tumors.
- 31. Verwendung der Polypeptid-Teilsequenzen Seq. ID No. 51–106, zur Herstellung eines Arzneimittels zur Be- handlung gegen den Blasen-tumor. 60
- 32. Arzneimittel, enthaltend mindestens eine Polypeptid-Teilsequenz Seq. ID No. 51–106.
- 33. Eine Nukleinsäure-Sequenz gemäß den Ansprüchen 1 bis 10, dadurch gekennzeichnet, daß es eine genomische Sequenz ist.
- 34. Eine Nukleinsäure-Sequenz gemäß den Ansprüchen 1 bis 10, dadurch gekennzeichnet, daß es eine mRNA-Se- quenz ist. 65
- 35. Genomische Gene, ihre Promotoren, Enhancer, Silencer, Exonstruktur, Intronstruktur und deren Spleißvarian- ten, erhältlich aus den cDNAs der Sequenzen Seq. ID No. 1 bis Seq. ID No. 50.
- 36. Verwendung der genomischen Gene gemäß Anspruch 33, zusammen mit geeigneten regulativen Elementen.

37. Verwendung gemäß Anspruch 34, dadurch gekennzeichnet, daß das regulative Element ein geeigneter Promotor und/oder Enhancer ist.

38. Eine Nukleinsäure-Sequenz gemäß den Ansprüchen 1 bis 7, dadurch gekennzeichnet, daß die Größe des Fragments eine Länge von mindestens 300 bis 3500 bp aufweist.

5

Hierzu 10 Seite(n) Zeichnungen

10

15

20

25

30

35

40

45

50

55

60

65

## Systematische Gen-Suche in der Incyte LifeSeq Datenbank

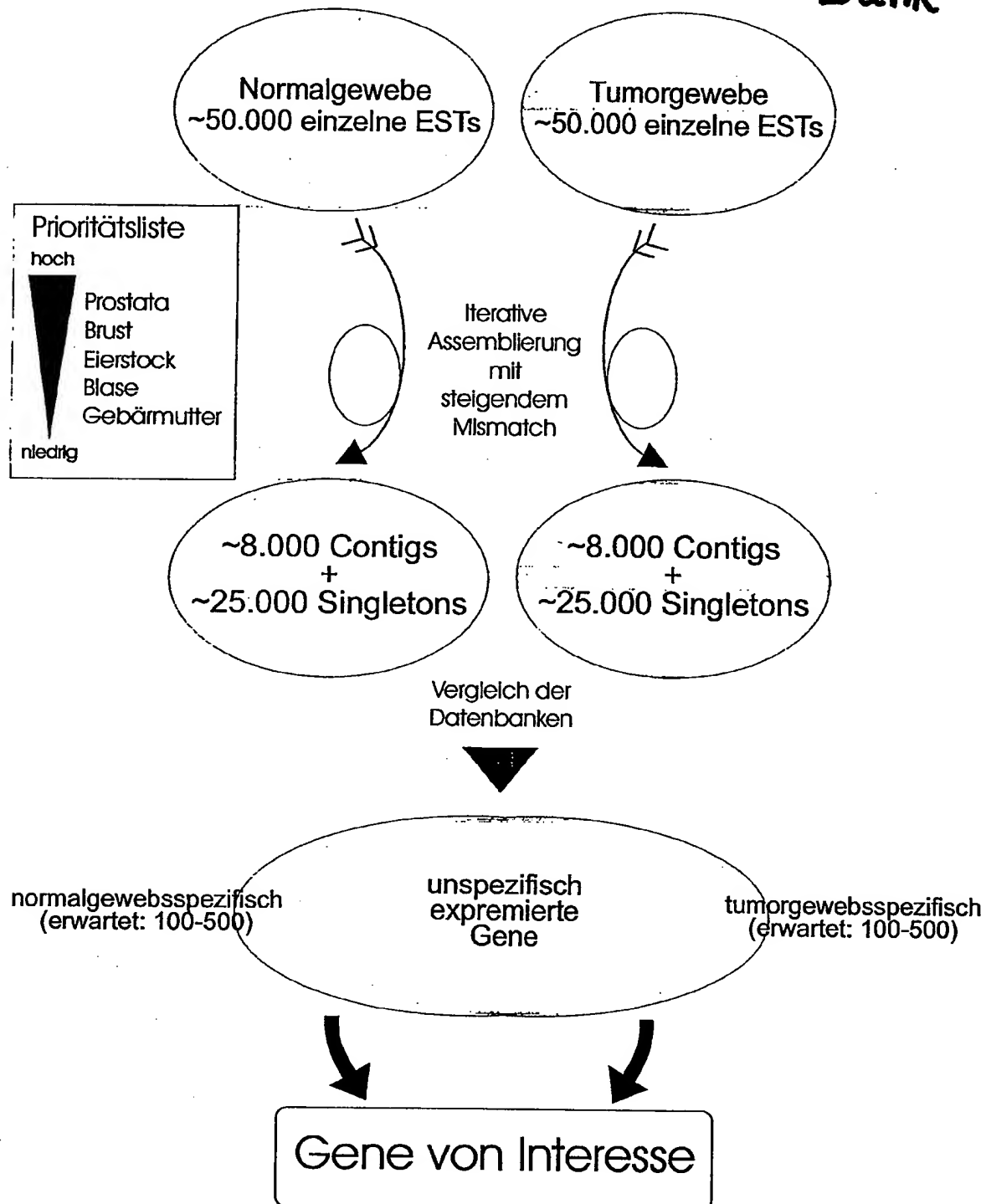


Fig. 1

# Prinzip der EST-Assemblierung

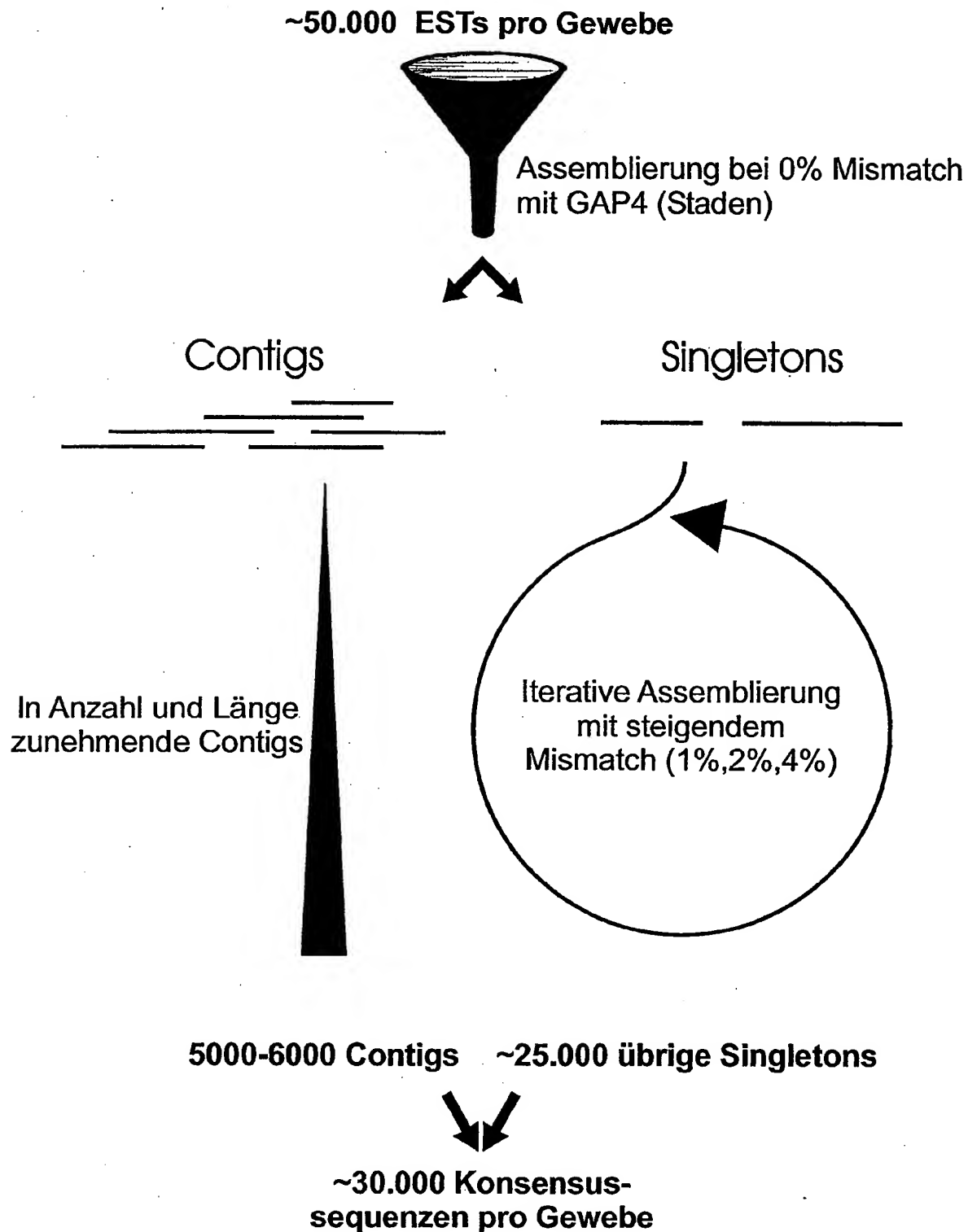


Fig. 2a

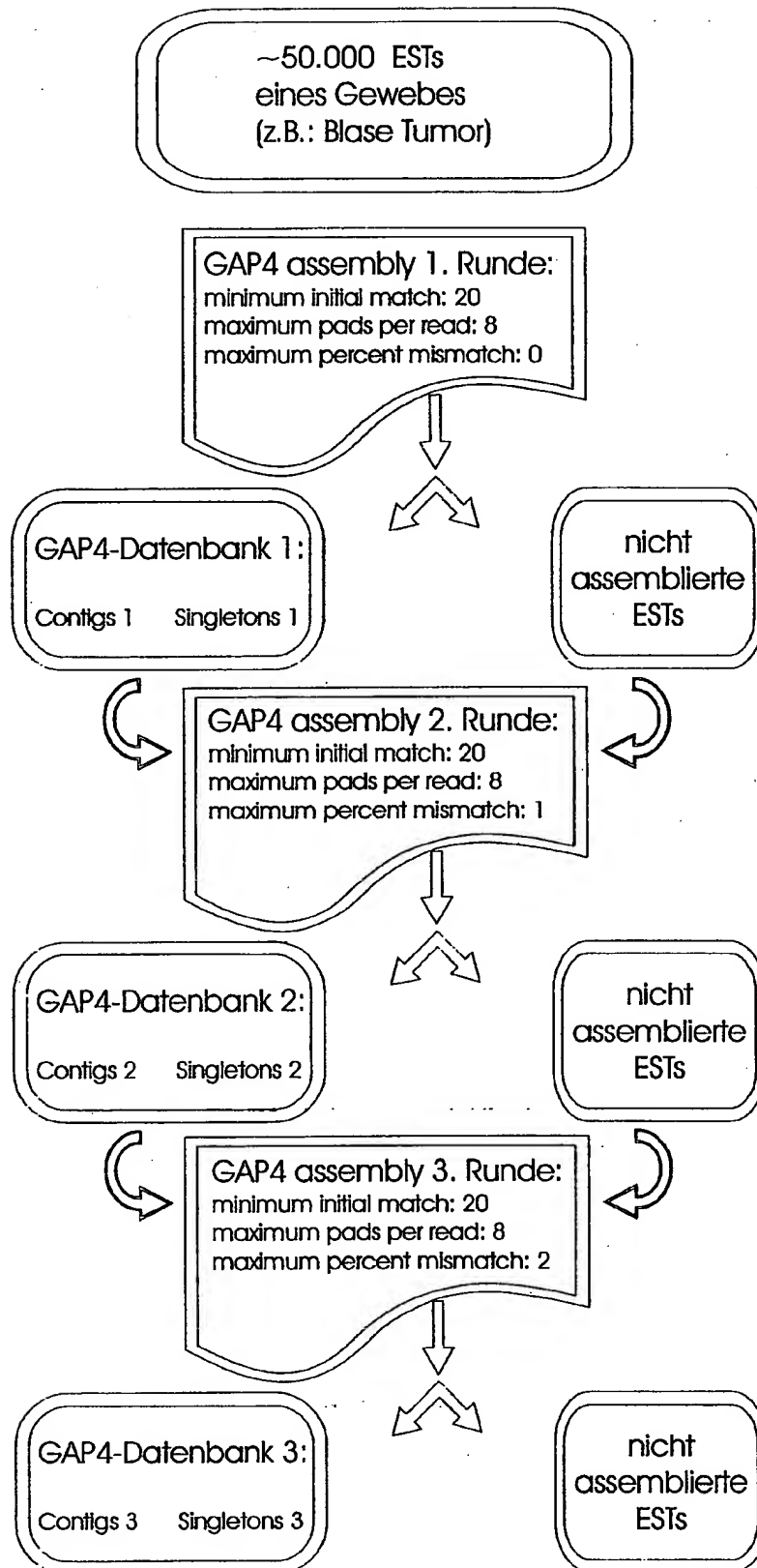


Fig. 2b1

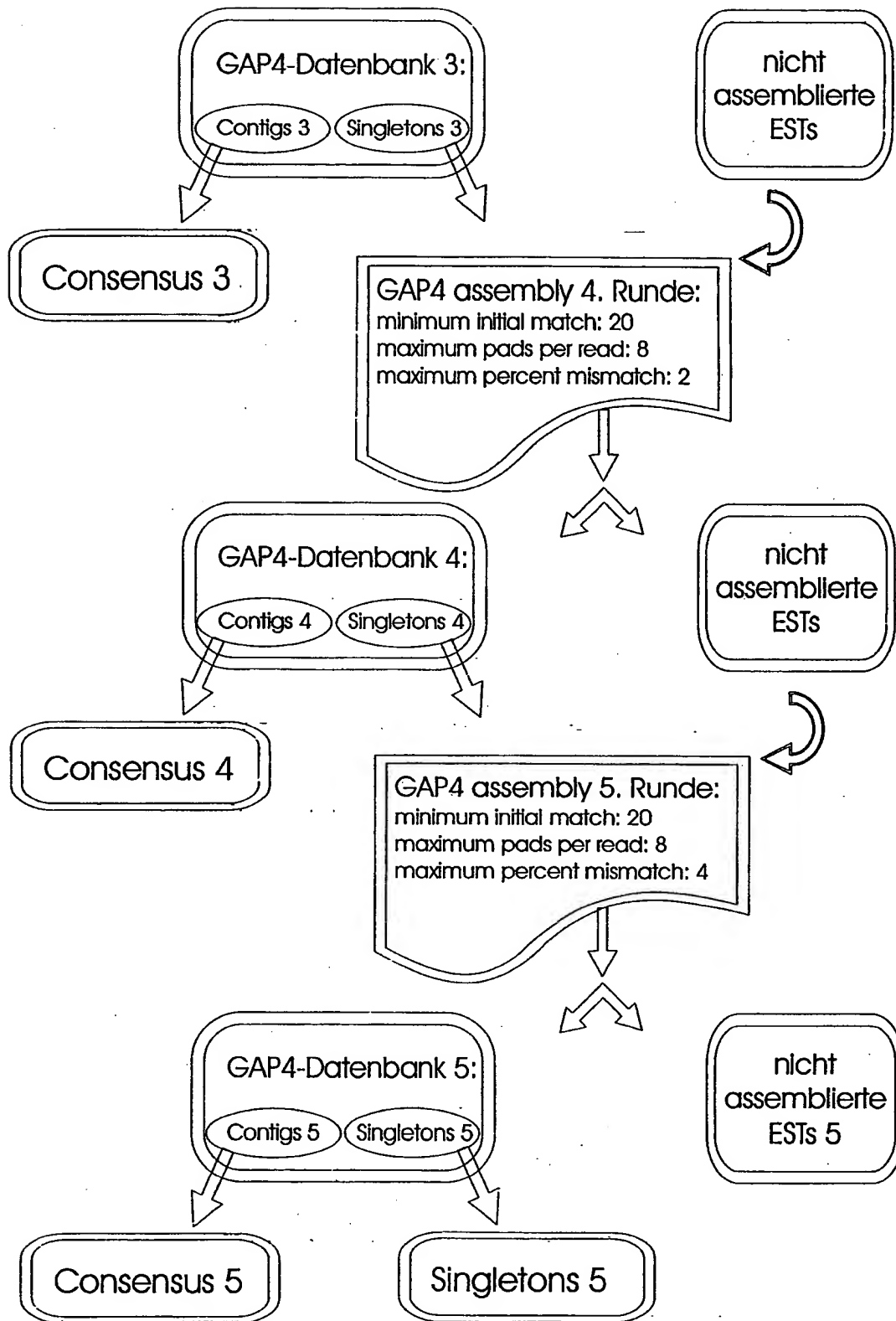


Fig. 2b2

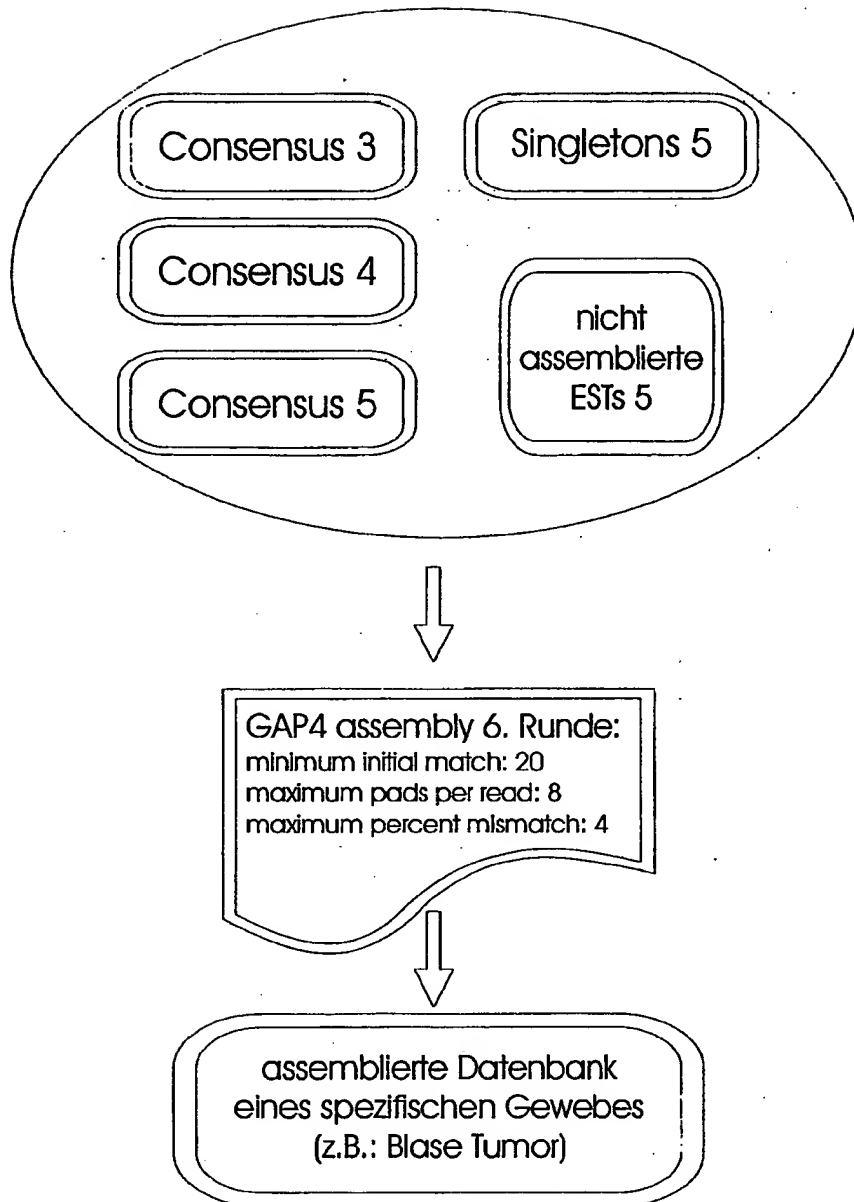


Fig. 2b3



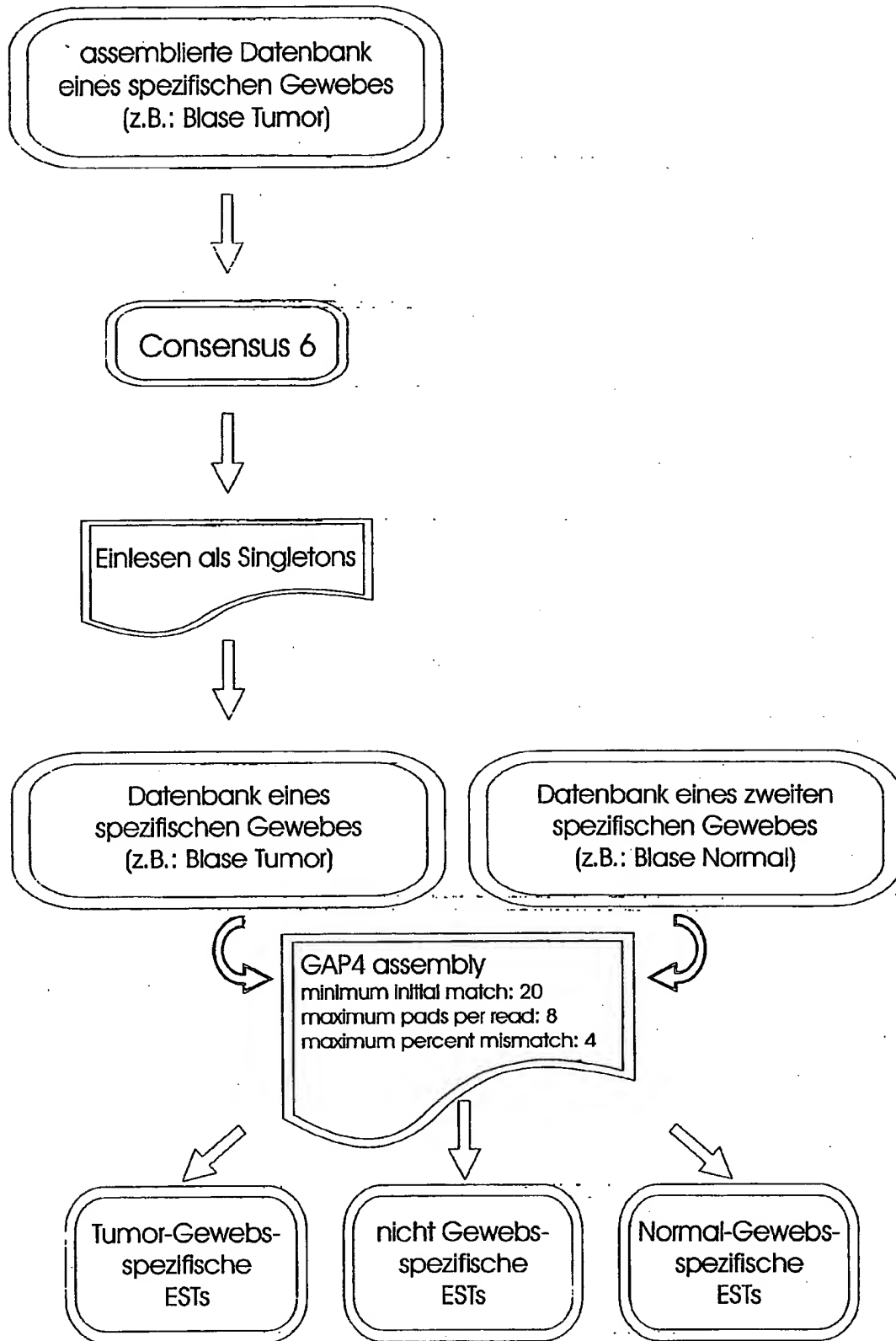


Fig. 2b4

# In silico Subtraktion der Genexpression in verschiedenen Geweben

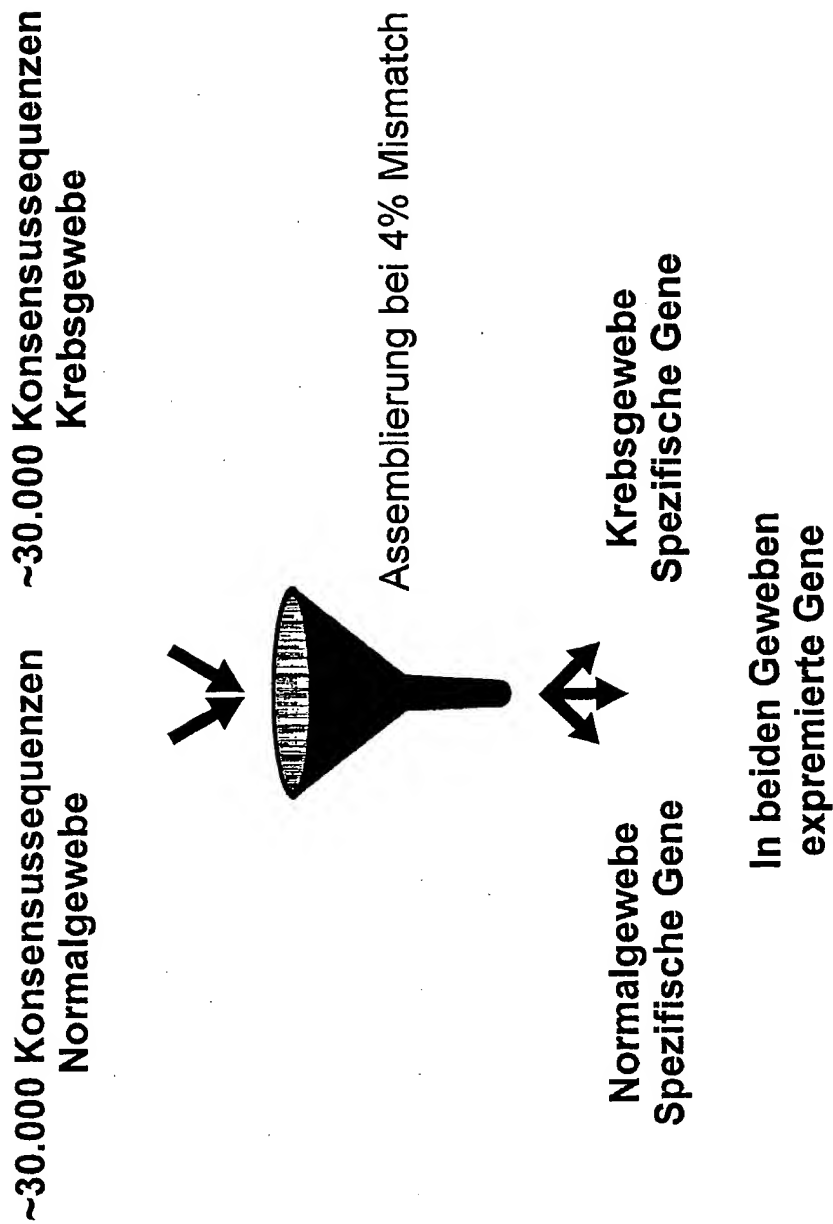


Fig. 3

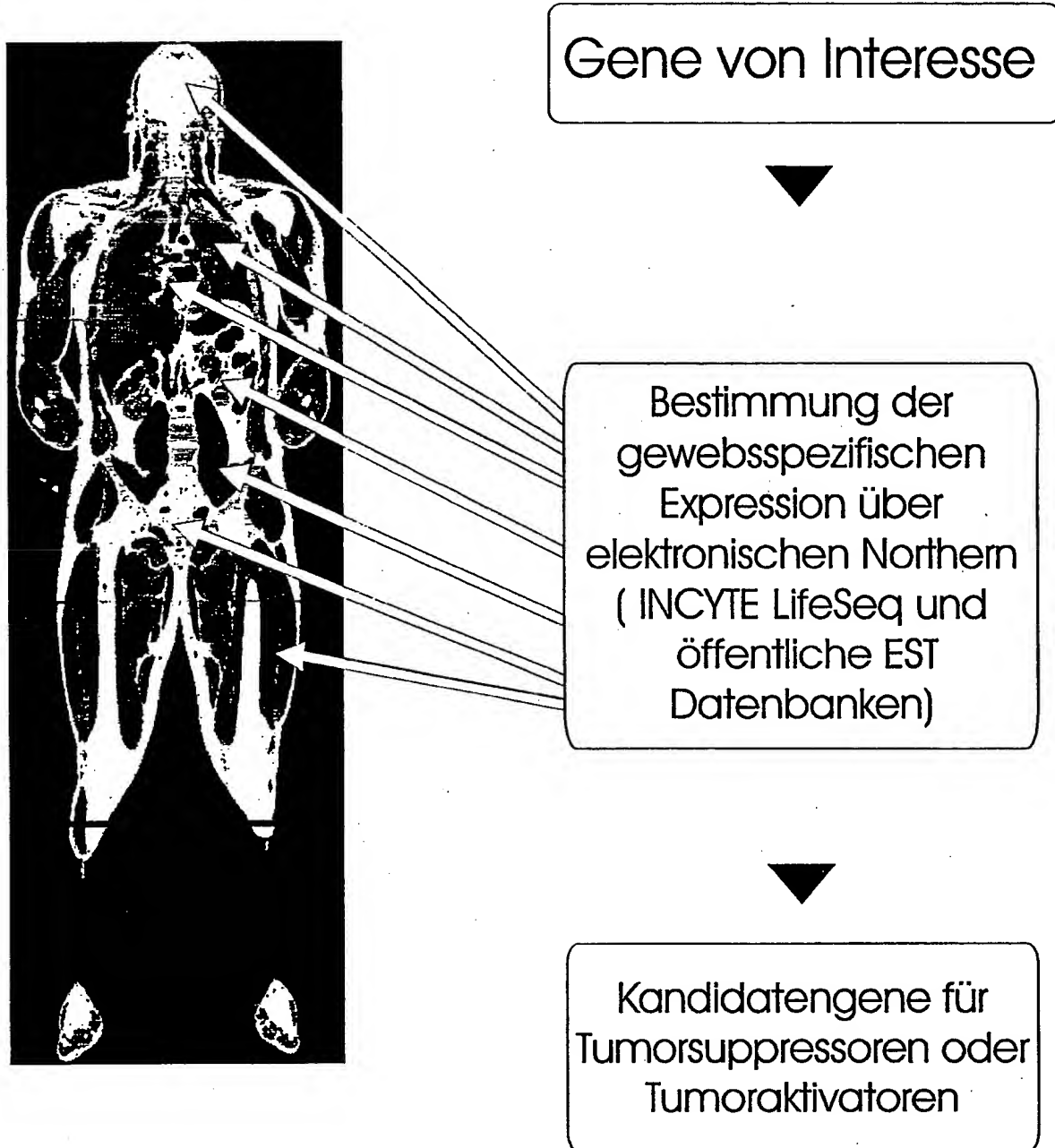


Fig. 4a

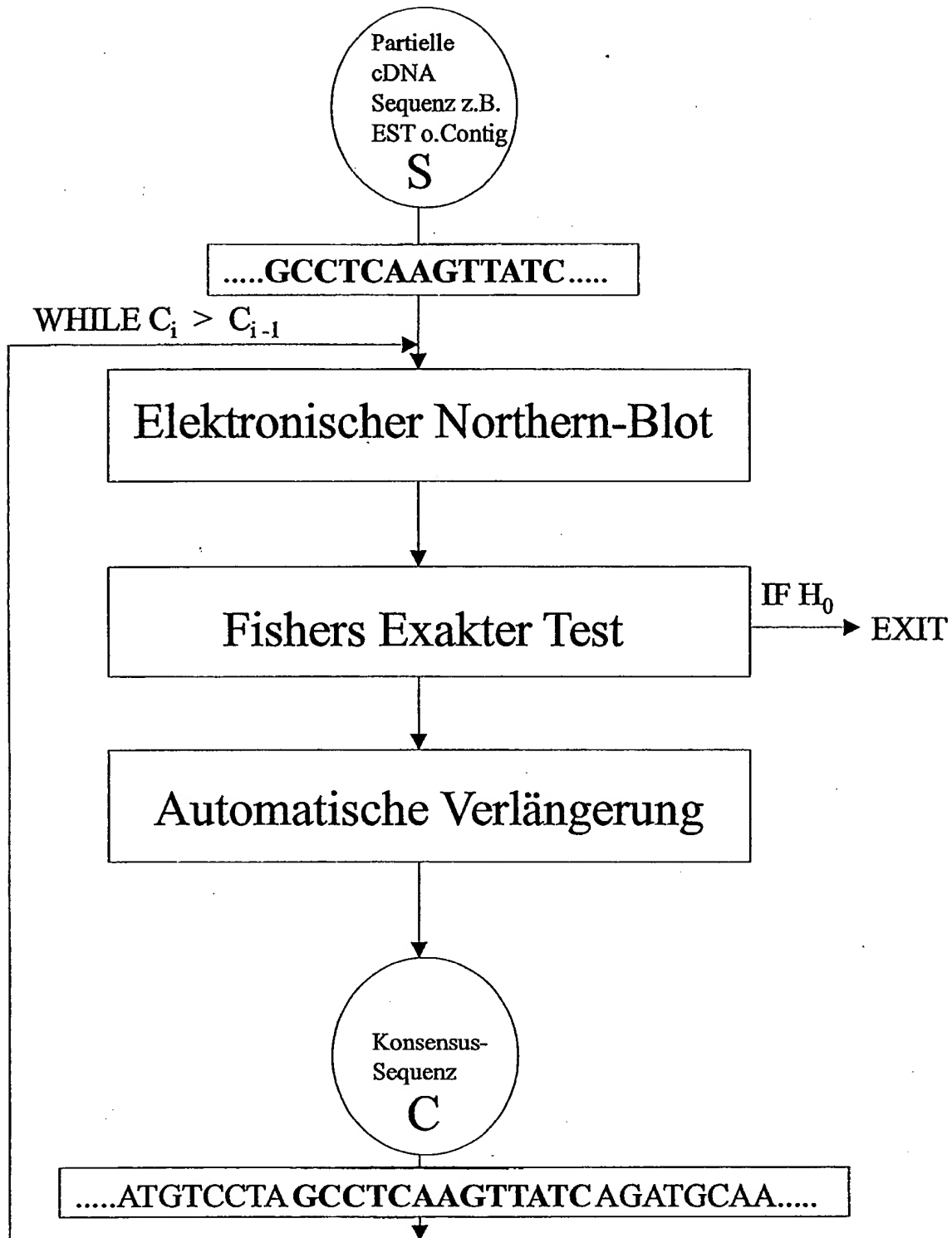


Fig. 4b

Isolieren von genomischen BAC und PAC Klonen



Chromosomale Klon-Lokalisation über FISH



Hybridisierungssignal



Sequenzierung von Klonen, die in Regionen lokalisiert sind, die chromosomale Deletionen in Prostata- und Brustkrebs aufweisen, führt zur Identifizierung von Kandidatengenen



Bestätigung der Kandidatengene durch Screening von Mutationen und/oder Deletionen in Krebsgeweben

Fig. 5

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 82:

ARAARGARRT SRAVTPTCAT PAGPMPCSL PPSLRCSLHS ACCSGDPASY RLWGAPLQPT 60  
LGVVPQASVP LLTDLAQWEP VLVPEAHPNA SLTMYVCTPV PHPDPPMALS RTPTRQISSS120  
DTDPPADGPS NPLCCCFHGP AFSTLNPVLR HLFPQEAFFA HPIYDLSQVW SVVSPAPSRG180  
QALRRAQ 187

(2) INFORMATION ÜBER SEQ ID NO: 83:

(A) LÄNGE: 241 Aminosäuren

(B) TYP: Protein

(C) STRANG: einzel

(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 83:

FFFFGLSNRC LLRAYAVLRL PFREPHECEA WPLPPGLQAP SLETPRNSRR LLSSSSTQST 60  
SSQPLLGPPE CLSPAGCGGH HGPDLAQVID GVGREGFLGE EVPEHRVKGG ECWAMETAAE120  
RVGGAICRRI CVTRADLPGG SPGEGHGRVR VGHARGADIHG QTCVRMCLRN QDRLPLGQVC180  
EEWHRGLGHH TQCGLQRGPP EPIAGRVPR AGRVQGAAGG WRQPAAGHGP RWRCTSRCHS240  
T 241

(2) INFORMATION ÜBER SEQ ID NO: 84:

(A) LÄNGE: 113 Aminosäuren

(B) TYP: Protein

(C) STRANG: einzel

(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 84:

MGWAGKASWG KRCLSTGLRV ENAGPWKQQQ RGLEGPSACG SVSLELICRV GVLERAMGGS 60  
GWGTGVQTYM VRLALGCASG TRTGSHWARS VRSGTEAWGT TPSVGOSGAP QSL 113

5 (2) INFORMATION ÜBER SEQ ID NO: 85:

- 10 (A) LÄNGE: 107 Aminosäuren  
(B) TYP: Protein  
(C) STRANG: einzel  
(D) TOPOLOGIE: linear

15 (ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

20 (vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

25 (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 85:

AFLSFLFSER FKASTTLFPP SLLNLICTKS FALVGVVETA LSLSTSVREC EPPWQVPVQG 60  
PAALHLGRVT GAPAVCPKAS PWPFGLSLGR FRTEHQGRQA FQGISIN 107

30 (2) INFORMATION ÜBER SEQ ID NO: 86:

- 35 (A) LÄNGE: 107 Aminosäuren  
(B) TYP: Protein  
(C) STRANG: einzel  
(D) TOPOLOGIE: linear

40 (ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

45 (vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

50 (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 86:

LRNRLWKVKL EEPDLISPTS KTPSEQQRPO HPPRTGDSIF MATPCGGRLT TSHHIPELS 60  
SSSGMTPSP PPSSSFSSFC LFVSELSCLS FFLRDSKPPR LCFPRPF 107

55 (2) INFORMATION ÜBER SEQ ID NO: 87:

- 60 (A) LÄNGE: 115 Aminosäuren  
(B) TYP: Protein  
(C) STRANG: einzel  
(D) TOPOLOGIE: linear

65 (ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:  
(A) ORGANISMUS: MENSCH

5

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 87:

IQKGRGKQSR GGFESLRKKE RQESSETKRQ KDEKEEEGGG GDGVIPLEEL SSGMMWWLVV 60  
NRPPQGVAMK MESPVRRGCC GRCCSDGVFD VGLMRSGSSS FTFQSRFLSQ VVQWL 115

10

(2) INFORMATION ÜBER SEQ ID NO: 88:

(A) LÄNGE: 124 Aminosäuren  
(B) TYP: Protein  
(C) STRANG: einzel  
(D) TOPOLOGIE: linear

15

(ii) MOLEKÜLTYP: ORF

20

(iii) HYPOTHETISCH: ja

25

(vi) HERKUNFT:  
(A) ORGANISMUS: MENSCH

30

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 88:

CSVRNLPRLR PKQGGEAFGH TAGAPVTRPR WRAAGPCTGT CQGGSHSRTL VLRLKAVSTT 60  
PTSANDLVQI KFRRDGGNRV VEALNLSEKR KDRKAQKQRD RRMKKKKREV VGTASSRWRS120  
SALG 124

35

(2) INFORMATION ÜBER SEQ ID NO: 89:

(A) LÄNGE: 198 Aminosäuren  
(B) TYP: Protein  
(C) STRANG: einzel  
(D) TOPOLOGIE: linear

40

(ii) MOLEKÜLTYP: ORF

45

(iii) HYPOTHETISCH: ja

50

(vi) HERKUNFT:  
(A) ORGANISMUS: MENSCH

55

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 89:

EGAGGEWRCP AAGGPRGEDG PPGLRLTERA GLTTRHLTGT ADPSQKHNL DPSSSKDERI 60  
QQTQPNREGA CSLNLHGLRY LCKPGVAPVL RVILRSCLEP NGFTSGSCRF SLGLSLILKW120  
VAGGRGWLRP LLALRLQWE QDTSPEFHFF SCPNHAHTIV QNQTSTFEKWL HGHPPGPRKL180  
HSKGLFTWQQ NPSPAVSP 198

60

65

(2) INFORMATION ÜBER SEQ ID NO: 90:



(A) LÄNGE: 124 Aminosäuren  
(B) TYP: Protein  
(C) STRANG: einzel  
(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 90:

LPPVEPPVLK GSCRKDMHRN EERERSENEV WRARPGPTAQ GSSPPPDAPF HPPPQCLLSP 60  
KAPASEVLGA HPPQCGQGGK GQVLDTAKCP EMTLLLTTHFF GPWQSPTCPQ HGAPGRTGRQ120  
EGGW 124

(2) INFORMATION ÜBER SEQ ID NO: 91:

(A) LÄNGE: 147 Aminosäuren  
(B) TYP: Protein  
(C) STRANG: einzel  
(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 91:

NSGEVSCSQL CSLRASRGRS HPLPPATHFK MRLRPRLKRQ LPEVNPFGRK HERRMTLRTG 60  
ATPGLHKYRR PWRLRLQAPS LLGCVCCIIS SLELLGSLRL CFWDGSAVPV RCLVVRPALS120  
VSLSPGGPSS PLGPPAAGHL HSPPAPS 147

(2) INFORMATION ÜBER SEQ ID NO: 92:

(A) LÄNGE: 374 Aminosäuren  
(B) TYP: Protein  
(C) STRANG: einzel  
(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

## (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 92:

SREAPESRRW AVWRSLESLP RHQLLCPLVG APPAPAMLSA LARPASAALR RSFSTSAQNN 60  
AKVAVLGASG GIGQPLSLLL KNSPLVSRLT LYDIAHTPGV AADLSHIETK AAVKGYLGPE120  
QLPDCLKGCD VVVIPAGVPR KPGMTRDDL FNTNATIVATL TAACAQHCPE AMICVIANPV180  
NSTIPITAEV FKKHGVYNPN KIFGVTTLDI VRANTFVAEL KGLDPA RVNV PVIGGHAGKT240  
IIP LISQCTP KVDFPQDQLT ALTGRIQEAG TEVVKAKAGA GSATLSMAYA GARFVFSLVD300  
AMNGKEGVVE CSFVKSQETE CTYFSTPLLL GKKGIEKNLG IGVSSFE EK MISDAIPELK360  
ASIKKGEDFV KTLK 374

## (2) INFORMATION ÜBER SEQ ID NO: 93:

(A) LÄNGE: 238 Aminosäuren

(B) TYP: Protein

(C) STRANG: einzel

(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

## (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 93:

LNEGTFHNTF LSIHCHKGE DKAGAGIRHG EGGRACSGFS LDHLRAGLLD PPSECCQLVL 60  
GKVHLGGALR DQGGDGLPSM ATNDRDVDSS WIQTLQLCNK GVGSDDVQGR HAEDFVGVVH120  
SMLENFCCD GDGGINRIGN DADHGFR AVL GTGSGQGGHN RGIGVEQVVP GHAWLSGDSS180  
RNNYHITTFQ AVRQLFRSEV AFHSGFGLDV AQICGHSGCV RDIIEGQAAH QGAVLQEK 238

## (2) INFORMATION ÜBER SEQ ID NO: 94:

(A) LÄNGE: 242 Aminosäuren

(B) TYP: Protein

(C) STRANG: einzel

(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

## (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 94:

EAGEEKAEEE GVAEEGVNK FSYP PSHREC CPAVEEEDDE EAVKKEA HRT STSALSPGSK 60  
PSTWVSCPGE EENQATEDKR TERSKGARKT SVSPRSSGSD PRSWEYRSGE ASEEEK EKAH120  
KETGKG E AAP GPOSSAP AQR PQLKSWWCQP SDEEEGEVKA LGAAEKDGEA ECPPCI PPSP180  
AFLKAWVYWP GEDTEEE EDE EDEDS DSGS DEEEGEAEAS SSTPATGVFL KSWVYQPGED240  
TQ 242

## (2) INFORMATION ÜBER SEQ ID NO: 95:

(A) LÄNGE: 237 Aminosäuren  
 (B) TYP: Protein  
 (C) STRANG: einzel  
 (D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:  
 (A) ORGANISMUS: MENSCH

## (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 95:

RPGCIGQERT LDSEDKEDDS EALGEAESD PHPSHPDORA HFRGWGYRPG KETEEEEAAE 60  
 DWGEAEPCPF RVAIYVPGEK PPPPWAPPRL PLRLQRRLLR PETPTHDPDP ETPLKARKVR120  
 FSEKVTVHFL AVWAGPAQAA RQGPWEQLAR DRSRFARRIT QAQEELSPCL TPAARARAWA180  
 RLRNPPLAPI PALTQTLPS SVPSSPVQTT PLSQAVATPS RSSAAAAAAL DLSGRRG 237

## (2) INFORMATION ÜBER SEQ ID NO: 96:

(A) LÄNGE: 890 Aminosäuren  
 (B) TYP: Protein  
 (C) STRANG: einzel  
 (D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:  
 (A) ORGANISMUS: MENSCH

## (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 96:

QDEHLITFFV PVFEPLPPQY FIRVVS DRWL SCETQLPVSE RHLILPEKYP PPTLLDLQP 60  
 LPVSALRNSA FESLYQDKFP FFNPIQTQVF NTVYNSDDNV FVGAPTGS GK TICAEFAILR120  
 MLLQSSEGR C VYITPMEALA EQVMDWYEK FQDRLNKKVV LLTGETSTDL KLLGKGNIII180  
 STPEKWDILS RRWKQRKNVQ NINLFVVDEV HLI GGENGPV LEVICSRMRY ISSQIERPIR240  
 IVALSSSLSN AKDVAHWLGC SATSTFNFHP NVRPVPLELH IQGFNISHTQ TRLLSMAKPV300  
 YHAITKHSPK KPVIVFVPSR KQTRLTAIDI LTTCAADIQR QRFLHCTEKD LIPYLEKLS D360  
 STLKETLLNG VGYLHEGLSP MERRLVEQLF SSGAIQVVVA SRSLCWGMNV AAHLVIIMDT420  
 QYNGKI HAY VDYPIYDVLQ MVGHANRPLQ DDEGR CVIMC QGSKKDFFKK FLYEPLPVES480  
 HLDHCMHDHF NAEIVTKTIE NKQDAVDYLT WFLYRRMTQ NPNYYNLQGI SHRHLSDHLS540  
 ELVEQTLSD L EQSKCISIED EMDVAPLNLG MIAAYYYINY TTIELFSMSL NAKTKVRGLI600  
 EIISNAAEYE NIPIRHEDN LLRLQAQKVP HKLNNPKFND PHVKTNLLQ AHL SRMQLSA660  
 ELQSDTEEIL SKAIRLIQAC VDVLSSNGWL SPALAAAMELA QMVTQAMWSK DSYLKQLPHF720  
 TSEHIK RCTD KGVESVFDIM EMEDEERNAL LQLTDSQIAD VARFCNRYPN IELSYEVVDK780  
 DSIRSGGPVV VLVQLEREEE VTGPVIAPLF PQKREEGWWV VIGDAKSNSL ISIKRLTLQQ840  
 KAKVKLDFVA PATGAHNYTL YFMSDAYMGC DQEKFSVDV KEAETDSDSD 890

## (2) INFORMATION ÜBER SEQ ID NO: 97:

- (A) LÄNGE: 281 Aminosäuren  
 (B) TYP: Protein  
 (C) STRANG: einzel  
 (D) TOPOLOGIE: linear

5

(ii) MOLEKÜLTYP: ORF

10

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:

15

(A) ORGANISMUS: MENSCH

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 97:

20

GDGSAEHGPR PLAAPLVTSR GAPASARPRG ALPGGSAPSA PHGQLPGRAQ PAPVSGPPPT 60  
 SGLCHFDPAA PWPLWPGPWQ LPPHPQDWPA QPDIPQDWVS FLRSFGQLTL CPRNGTVTGK120  
 WRGSHVVGILL TTLNFGDGPD RNKTRTFQAT VLGSQMGLKG SSAGQLVLIT ARVTTERTAG180  
 TCLYFSAVPG ILPSSQPPIS CSEEGAGNAT LSPRMGEECV SVWSHEGLVL TKLLTSEELA240  
 LCGSRLLVLG SFLLLFCGLL CCVTAMCFHP RRESHWSRTR L 281

25

(2) INFORMATION ÜBER SEQ ID NO: 98:

- (A) LÄNGE: 206 Aminosäuren  
 (B) TYP: Protein  
 (C) STRANG: einzel  
 (D) TOPOLOGIE: linear

30

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

40

(vi) HERKUNFT:

45

(A) ORGANISMUS: MENSCH

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 98:

50

RLEPRSVTRS RRAVSRLSAR PGKVSAMAF LASGPYLTHQ QKVLRLYKRA LRHLESWCVQ 60  
 RDKYRYFACL MRARFEEHKN EKDMAKATQL LKEAEEEFWY RQHPQPYIFP DSPGGTSYER120  
 YDCYKVPWC LDDWHPSEKA MYPDYFAKRE QWKKLRRESW EREVKQLQEE TPPGGPLTEA180  
 LPPARKEGDL PPLWYIIVTR PRERPM 206

55

(2) INFORMATION ÜBER SEQ ID NO: 99:

- (A) LÄNGE: 139 Aminosäuren  
 (B) TYP: Protein  
 (C) STRANG: einzel  
 (D) TOPOLOGIE: linear

60

(ii) MOLEKÜLTYP: ORF

65

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 99:

PLVPSFPSAV SSTVLSWQSN QDTLPSQKDA SHLSTILGPC SNRISHRRCP QESQGRMAV 60  
DADGTRILPR PPSAAGWPSP YPFHSYVLQT GLSSNKQSIG ICLSGRTTTR GGVAPAYKAA120  
TPFADGSGRV PTPRTPLRR 139

(2) INFORMATION ÜBER SEQ ID NO: 100:

(A) LÄNGE: 79 Aminosäuren

(B) TYP: Protein

(C) STRANG: einzel

(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 100:

APFWDLVAIV SLIGGAPRRV REDVWLWMLT VPEFFLGLLQ QLGGLRHILF ILMFFKPGSH60  
QTSKVSFVFS LDAPRLEVA 79

(2) INFORMATION ÜBER SEQ ID NO: 101:

(A) LÄNGE: 89 Aminosäuren

(B) TYP: Protein

(C) STRANG: einzel

(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 101:

VGGACAVALP QAAAMAGQED PVQREIHQDW ANREYIEIIT SSIKKIADFL NSFDMSCRSR60  
LATLNEKLTA LERRIEYIEA RVTKGETLT 89

(2) INFORMATION ÜBER SEQ ID NO: 102:

(A) LÄNGE: 88 Aminosäuren

- (B) TYP: Protein
- (C) STRANG: einzel
- (D) TOPOLOGIE: linear

5

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

10

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

15

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 102:

NSAVLLLGSC FTQHRPHGKG PSSQLLPFS LKSNRAYSCF SFFKSVAFGL CHLGCGVVCG60  
KKFRGTVGND VRHFTFSVTF YTSTCQCI 88

20

(2) INFORMATION ÜBER SEQ ID NO: 103:

25

(A) LÄNGE: 89 Aminosäuren

(B) TYP: Protein

(C) STRANG: einzel

(D) TOPOLOGIE: linear

30

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

35

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

40

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 103:

HFIHLLVNVF ETFTAKSLGL FVKVLLTSIF LSLSLKLSLK FSLPLHCFCE QSLSPPHLWW60  
EVRGSPGQDT HALAMWLPEN VVANPPVSC 89

45

(2) INFORMATION ÜBER SEQ ID NO: 104:

50

(A) LÄNGE: 240 Aminosäuren

(B) TYP: Protein

(C) STRANG: einzel

(D) TOPOLOGIE: linear

55

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

60

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

65

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 104:

REQILFIEIR DTAKGGETEQ PPSLSPLHGG RMPMEGEGIQ SLARETQSHR GRRQGWATW 60  
 5 VTRCRESLNR GGAGAGKRAG ALAHHVFLAL IEPNLAAREA SEEEVKACSD ETVVADLLVK120  
 VVYVLGAILK IFLREGNVLN QHSGMDIEKY SEHYQHDHSP GAEDDAAGGQ LRPTAQERRH180  
 KEGSRGSPRC KRARKAVGES PGCPRRGAEG AWPRPQPSHP GAAGGKGWAP LRTFAGDGAV240

(2) INFORMATION ÜBER SEQ ID NO: 105:

(A) LÄNGE: 136 Aminosäuren

(B) TYP: Protein

(C) STRANG: einzel

(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 105:

RLYMFWGPFS KSFSVKGTSS INTAEWTLKN IPSIINMITP QALRTMPQAA SFGPQHRSDA 60  
 IKKGVGDRRG ASGLGKRWEK AQDALAGGQR GRGPGLNHPI RGRQAEKAGL LSGLSRETAP120  
 SETKTAPGET FLDLCS 136

(2) INFORMATION ÜBER SEQ ID NO: 106:

(A) LÄNGE: 173 Aminosäuren

(B) TYP: Protein

(C) STRANG: einzel

(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 106:

LQSRKVSPPG AVLVS DGAVS RESPERSPAF SACRPRMGWL RPGPRPLCPP ARASWAFSHR 60  
 FPSPLAPRRS PTPFFMASLL CCGPKLAACG IVLSAWGVIM LIMLGIFFNH HSAVLIEDVP120  
 FTEKDFENGQ QNIYNLYEQV SYNCFIAAGL YLLLGGFSFC QVRLNKRKEY MVR 173

Patentansprüche

1. Eine Nukleinsäure-Sequenz, die ein Genprodukt oder ein Teil davon kodiert, umfassend
  - a) eine Nukleinsäure-Sequenz, ausgewählt aus der Gruppe Seq. ID No. 2-5, 7-13, 16, 18, 20, 23, 26-27, 31-32, 36, 45.

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Ashkenazi et al.	Group Art Unit: 1647
Serial No.: 09/903,925	Examiner: Fozia Hamid
Filed: July 11, 2001	<b>CERTIFICATE OF MAILING</b> I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Assistant Commissioner of Patents, Washington, D.C. 20231 on
For: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS	Date

DECLARATION OF AUDREY D. GODDARD, Ph.D UNDER 37 C.F.R. § 1.132

Assistant Commissioner of Patents  
Washington, D.C. 20231

Sir:

I, Audrey D. Goddard, Ph.D. do hereby declare and say as follows:

1. I am a Senior Clinical Scientist at the Experimental Medicine/BioOncology, Medical Affairs Department of Genentech, Inc., South San Francisco, California 94080.
2. Between 1993 and 2001, I headed the DNA Sequencing Laboratory at the Molecular Biology Department of Genentech, Inc. During this time, my responsibilities included the identification and characterization of genes contributing to the oncogenic process, and determination of the chromosomal localization of novel genes.
3. My scientific Curriculum Vitae, including my list of publications, is attached to and forms part of this Declaration (Exhibit A).



Serial No.: \*

Filed: \*

4. I am familiar with a variety of techniques known in the art for detecting and quantifying the amplification of oncogenes in cancer, including the quantitative TaqMan PCR (i.e., "gene amplification") assay described in the above captioned patent application.

5. The TaqMan PCR assay is described, for example, in the following scientific publications: Higuchi *et al.*, Biotechnology 10:413-417 (1992) (Exhibit B); Livak *et al.*, PCR Methods Appl., 4:357-362 (1995) (Exhibit C) and Heid *et al.*, Genome Res. 6:986-994 (1996) (Exhibit D). Briefly, the assay is based on the principle that successful PCR yields a fluorescent signal due to Taq DNA polymerase-mediated exonuclease digestion of a fluorescently labeled oligonucleotide that is homologous to a sequence between two PCR primers. The extent of digestion depends directly on the amount of PCR, and can be quantified accurately by measuring the increment in fluorescence that results from decreased energy transfer. This is an extremely sensitive technique, which allows detection in the exponential phase of the PCR reaction and, as a result, leads to accurate determination of gene copy number.

6. The quantitative fluorescent TaqMan PCR assay has been extensively and successfully used to characterize genes involved in cancer development and progression. Amplification of protooncogenes has been studied in a variety of human tumors, and is widely considered as having etiological, diagnostic and prognostic significance. This use of the quantitative TaqMan PCR assay is exemplified by the following scientific publications: Pennica *et al.*, Proc. Natl. Acad. Sci. USA 95(25):14717-14722 (1998) (Exhibit E); Pitti *et al.*, Nature 396(6712):699-703 (1998) (Exhibit F) and Bieche *et al.*, Int. J. Cancer 78:661-666 (1998) (Exhibit G), the first two of which I am co-author. In particular, Pennica *et al.* have used the quantitative TaqMan PCR assay to study relative gene amplification of WISP and c-myc in various cell lines, colorectal tumors and normal mucosa. Pitti *et al.* studied the genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer, using the quantitative TaqMan PCR assay. Bieche *et al.* used the assay to study gene amplification in breast cancer.

Serial No.: \*

Filed: \*

7. It is my personal experience that the quantitative TaqMan PCR technique is technically sensitive enough to detect at least a 2-fold increase in gene copy number relative to control. It is further my considered scientific opinion that an at least 2-fold increase in gene copy number in a tumor tissue sample relative to a normal (i.e., non-tumor) sample is significant and useful in that the detected increase in gene copy number in the tumor sample relative to the normal sample serves as a basis for using relative gene copy number as quantitated by the TaqMan PCR technique as a diagnostic marker for the presence or absence of tumor in a tissue sample of unknown pathology. Accordingly, a gene identified as being amplified at least 2-fold by the quantitative TaqMan PCR assay in a tumor sample relative to a normal sample is useful as a marker for the diagnosis of cancer, for monitoring cancer development and/or for measuring the efficacy of cancer therapy.

8. I declare further that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true. I declare that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Jan. 16, 2003  
Date

Audrey D. Goddard  
Audrey D. Goddard, Ph.D.

**AUDREY D. GODDARD, Ph.D.**

Genentech, Inc.  
1 DNA Way  
South San Francisco, CA, 94080  
650.225.6429  
goddarda@gene.com

110 Congo St.  
San Francisco, CA, 94131  
415.841.9154  
415.819.2247 (mobile)  
agoddard@pacbell.net

**PROFESSIONAL EXPERIENCE**

**Genentech, Inc.**  
**South San Francisco, CA**

**1993-present**

**2001 - present      Senior Clinical Scientist**  
Experimental Medicine / BioOncology, Medical Affairs

**Responsibilities:**

- *Companion diagnostic oncology products*
- *Acquisition of clinical samples from Genentech's clinical trials for translational research*
- *Translational research using clinical specimen and data for drug development and diagnostics*
- *Member of Development Science Review Committee, Diagnostic Oversight Team, 21 CFR Part 11 Subteam*

**Interests:**

- *Ethical and legal implications of experiments with clinical specimens and data*
- *Application of pharmacogenomics in clinical trials*

**1998 - 2001      Senior Scientist**

Head of the DNA Sequencing Laboratory, Molecular Biology Department, Research

**Responsibilities:**

- *Management of a laboratory of up to nineteen –including postdoctoral fellow, associate scientist, senior research associate and research assistants/associate levels*
- *Management of a \$750K budget*
- *DNA sequencing core facility supporting a 350+ person research facility.*
- *DNA sequencing for high throughput gene discovery, - ESTs, cDNAs, and constructs*
- *Genomic sequence analysis and gene identification*
- *DNA sequence and primary protein analysis*

**Research:**

- *Chromosomal localization of novel genes*
- *Identification and characterization of genes contributing to the oncogenic process*
- *Identification and characterization of genes contributing to inflammatory diseases*
- *Design and development of schemes for high throughput genomic DNA sequence analysis*
- *Candidate gene prediction and evaluation*

**1993 - 1998            Scientist**

Head of the DNA Sequencing Laboratory, Molecular Biology Department, Research

**Responsibilities**

- *DNA sequencing core facility supporting a 350+ person research facility*
- *Assumed responsibility for a pre-existing team of five technicians and expanded the group into fifteen, introducing a level of middle management and additional areas of research*
- *Participated in the development of the basic plan for high throughput secreted protein discovery program – sequencing strategies, data analysis and tracking, database design*
- *High throughput EST and cDNA sequencing for new gene identification.*
- *Design and implementation of analysis tools required for high throughput gene identification.*
- *Chromosomal localization of genes encoding novel secreted proteins.*

**Research:**

- *Genomic sequence scanning for new gene discovery.*
- *Development of signal peptide selection methods.*
- *Evaluation of candidate disease genes.*
- *Growth hormone receptor gene SNPs in children with Idiopathic short stature*

**Imperial Cancer Research Fund  
London, UK with Dr. Ellen Solomon**

**1989-1992**

**6/89 –12/92 Postdoctoral Fellow**

- Cloning and characterization of the genes fused at the acute promyelocytic leukemia translocation breakpoints on chromosomes 17 and 15.
- Prepared a successfully funded European Union multi-center grant application

**McMaster University  
Hamilton, Ontario, Canada with Dr. G. D. Sweeney**

**1983**

**5/83 – 8/83: NSERC Summer Student**

- *In vitro* metabolism of  $\beta$ -naphthoflavone in C57Bl/6J and DBA mice

**EDUCATION**

**Ph.D.**

"Phenotypic and genotypic effects of mutations in the human retinoblastoma gene."

**Supervisor:** Dr. R. A. Phillips

University of Toronto  
Toronto, Ontario, Canada.  
Department of Medical  
Biophysics.

**1989**

**Honours B.Sc**

"The *in vitro* metabolism of the cytochrome P-448 inducer  $\beta$ -naphthoflavone in C57BL/6J mice."

**Supervisor:** Dr. G. D. Sweeney

McMaster University,  
Hamilton, Ontario, Canada.  
Department of Biochemistry

**1983**

## ACADEMIC AWARDS

Imperial Cancer Research Fund Postdoctoral Fellowship	1989-1992
Medical Research Council Studentship	1983-1988
NSERC Undergraduate Summer Research Award	1983
Society of Chemical Industry Merit Award (Hons. Biochem.)	1983
Dr. Harry Lyman Hooker Scholarship	1981-1983
J.L.W. Gill Scholarship	1981-1982
Business and Professional Women's Club Scholarship	1980-1981
Wyerhauser Foundation Scholarship	1979-1980

## INVITED PRESENTATIONS

Genentech's gene discovery pipeline: High throughput identification, cloning and characterization of novel genes. Functional Genomics: From Genome to Function, Litchfield Park, AZ, USA. October 2000

High throughput identification, cloning and characterization of novel genes. G2K:Back to Science, Advances in Genome Biology and Technology I. Marco Island, FL, USA. February 2000

Quality control in DNA Sequencing: The use of Phred and Phrap. Bay Area Sequencing Users Meeting, Berkeley, CA, USA. April 1999

High throughput secreted protein identification and cloning. Tenth International Genome Sequencing and Analysis Conference, Miami, FL, USA. September 1998

The evolution of DNA sequencing: The Genentech perspective. Bay Area Sequencing Users Meeting, Berkeley, CA, USA. May 1998

Partial Growth Hormone Insensitivity: The role of GH-receptor mutations in Idiopathic Short Stature. Tenth Annual National Cooperative Growth Study Investigators Meeting, San Francisco, CA, USA. October, 1996

Growth hormone (GH) receptor defects are present in selected children with non-GH-deficient short stature: A molecular basis for partial GH-insensitivity. 76<sup>th</sup> Annual Meeting of The Endocrine Society, Anaheim, CA, USA. June 1994

A previously uncharacterized gene, myl, is fused to the retinoic acid receptor alpha gene in acute promyelocytic leukemia. XV International Association for Comparative Research on Leukemia and Related Disease, Padua, Italy. October 1991

## PATENTS

Goddard A, Godowski PJ, Gurney AL. NL2 Tie ligand homologue polypeptide. Patent Number: 6,455,496. Date of Patent: Sept. 24, 2002.

**Goddard A**, Godowski PJ and Gurney AL. NL3 Tie ligand homologue nucleic acids. Patent Number: 6,426,218. Date of Patent: July 30, 2002.

Godowski P, Gurney A, Hillan KJ, Botstein D, **Goddard A**, Roy M, Ferrara N, Tumas D, Schwall R. NL4 Tie ligand homologue nucleic acid. Patent Number: 6,4137,770. Date of Patent: July 2, 2002.

Ashkenazi A, Fong S, **Goddard A**, Gurney AL, Napier MA, Tumas D, Wood WI. Nucleic acid encoding A-33 related antigen poly peptides. Patent Number: 6,410,708. Date of Patent: Jun. 25, 2002.

Botstein DA, Cohen RL, **Goddard AD**, Gurney AL, Hillan KJ, Lawrence DA, Levine AJ, Pennica D, Roy MA and Wood WI. WISP polypeptides and nucleic acids encoding same. Patent Number: 6,387,657. Date of Patent: May 14, 2002.

**Goddard A**, Godowski PJ and Gurney AL. Tie ligands. Patent Number: 6,372,491. Date of Patent: April 16, 2002.

Godowski PJ, Gurney AL, **Goddard A** and Hillan K. TIE ligand homologue antibody. Patent Number: 6,350,450. Date of Patent: Feb. 26, 2002.

Fong S, Ferrara N, **Goddard A**, Godowski PJ, Gurney AL, Hillan K and Williams PM. Tie receptor tyrosine kinase ligand homologues. Patent Number: 6,348,351. Date of Patent: Feb. 19, 2002.

**Goddard A**, Godowski PJ and Gurney AL. Ligand homologues. Patent Number: 6,348,350. Date of Patent: Feb. 19, 2002.

Attie KM, Carlsson LMS, Gesundheit N and **Goddard A**. Treatment of partial growth hormone insensitivity syndrome. Patent Number: 6,207,640. Date of Patent: March 27, 2001.

Fong S, Ferrara N, **Goddard A**, Godowski PJ, Gurney AL, Hillan K and Williams PM. Nucleic acids encoding NL-3. Patent Number: 6,074,873. Date of Patent: June 13, 2000

Attie K, Carlsson LMS, Gesundheit N and **Goddard A**. Treatment of partial growth hormone insensitivity syndrome. Patent Number: 5,824,642. Date of Patent: October 20, 1998

Attie K, Carlsson LMS, Gesundheit N and **Goddard A**. Treatment of partial growth hormone insensitivity syndrome. Patent Number: 5,646,113. Date of Patent: July 8, 1997

Multiple additional provisional applications filed

## PUBLICATIONS

Seshasayee D, Dowd P, Gu Q, Erickson S, **Goddard AD**. Comparative sequence analysis of the *HER2* locus in mouse and man. Manuscript in preparation.

Abuzzahab MJ, **Goddard A**, Grigorescu F, Lautier C, Smith RJ and Chernausk SD. Human IGF-1 receptor mutations resulting in pre- and post-natal growth retardation. Manuscript in preparation.

Aggarwal S, Xie, M-H, Foster J, Frantz G, Stinson J, Corpuz RT, Simmons L, Hillan K, Yansura DG, Vandlen RL, **Goddard AD** and Gurney AL. FHFR, a novel receptor for the fibroblast growth factors. Manuscript submitted.

Adams SH, Chui C, Schilbach SL, Yu XX, **Goddard AD**, Grimaldi JC, Lee J, Dowd P, Colman S., Lewin DA. (2001) BFIT, a unique acyl-CoA thioesterase induced in thermogenic brown adipose tissue: Cloning, organization of the human gene, and assessment of a potential link to obesity. *Biochemical Journal* **360**: 135-142.

Lee J, Ho WH, Maruoka M, Corpuz RT, Baldwin DT, Foster JS, **Goddard AD**, Yansura DG, Vandlen RL, Wood WI, Gurney AL. (2001) IL-17E, a novel proinflammatory ligand for the IL-17 receptor homolog IL-17Rh1. *Journal of Biological Chemistry* **276**(2): 1660-1664.

Xie M-H, Aggarwal S, Ho W-H, Foster J, Zhang Z, Stinson J, Wood WI, **Goddard AD** and Gurney AL. (2000) Interleukin (IL)-22, a novel human cytokine that signals through the interferon-receptor related proteins CRF2-4 and IL-22R. *Journal of Biological Chemistry* **275**: 31335-31339.

Weiss GA, Watanabe CK, Zhong A, **Goddard A** and Sidhu SS. (2000) Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc. Natl. Acad. Sci. USA* **97**: 8950-8954.

Guo S, Yamaguchi Y, Schilbach S, Wada T.; Lee J, **Goddard A**, French D, Handa H, Rosenthal A. (2000) A regulator of transcriptional elongation controls vertebrate neuronal development. *Nature* **408**: 366-369.

Yan M, Wang L-C, Hymowitz SG, Schilbach S, Lee J, **Goddard A**, de Vos AM, Gao WQ, Dixit VM. (2000) Two-amino acid molecular switch in an epithelial morphogen that regulates binding to two distinct receptors. *Science* **290**: 523-527.

Sehl PD, Tai JTN, Hillan KJ, Brown LA, **Goddard A**, Yang R, Jin H and Lowe DG. (2000) Application of cDNA microarrays in determining molecular phenotype in cardiac growth, development, and response to injury. *Circulation* **101**: 1990-1999.

Guo S, Brush J, Teraoka H, **Goddard A**, Wilson SW, Mullins MC and Rosenthal A. (1999) Development of noradrenergic neurons in the zebrafish hindbrain requires BMP, FGF8, and the homeodomain protein soulless/Phox2A. *Neuron* **24**: 555-566.

Stone D, Murone, M, Luoh, S, Ye W, Armanini P, Gurney A, Phillips HS, Brush, J, **Goddard A**, de Sauvage FJ and Rosenthal A. (1999) Characterization of the human suppressor of fused; a negative regulator of the zinc-finger transcription factor Gli. *J. Cell Sci.* **112**: 4437-4448.

Xie M-H, Holcomb I, Deuel B, Dowd P, Huang A, Vagts A, Foster J, Liang J, Brush J, Gu Q, Hillan K, **Goddard A** and Gurney, A.L. (1999) FGF-19, a novel fibroblast growth factor with unique specificity for FGFR4. *Cytokine* **11**: 729-735.

- Yan M, Lee J, Schilbach S, **Goddard A** and Dixit V. (1999) mE10, a novel caspase recruitment domain-containing proapoptotic molecule. *J. Biol. Chem.* **274**(15): 10287-10292.
- Gurney AL, Marsters SA, Huang RM, Pitti RM, Mark DT, Baldwin DT, Gray AM, Dowd P, Brush J, Heldens S, Schow P, **Goddard AD**, Wood WI, Baker KP, Godowski PJ and Ashkenazi A. (1999) Identification of a new member of the tumor necrosis factor family and its receptor, a human ortholog of mouse GITR. *Current Biology* **9**(4): 215-218.
- Ridgway JBB, Ng E, Kern JA, Lee J, Brush J, **Goddard A** and Carter P. (1999) Identification of a human anti-CD55 single-chain Fv by subtractive panning of a phage library using tumor and nontumor cell lines. *Cancer Research* **59**: 2718-2723.
- Pitti RM, Marsters SA, Lawrence DA, Roy M, Kischkel FC, Dowd P, Huang A, Donahue CJ, Sherwood SW, Baldwin DT, Godowski PJ, Wood WI, Gurney AL, Hillan KJ, Cohen RL, **Goddard AD**, Botstein D and Ashkenazi A. (1998) Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer. *Nature* **396**(6712): 699-703.
- Pennica D, Swanson TA, Welsh JW, Roy MA, Lawrence DA, Lee J, Brush J, Taneyhill LA, Deuel B, Lew M, Watanabe C, Cohen RL, Melhem MF, Finley GG, Quirke P, **Goddard AD**, Hillan KJ, Gurney AL, Botstein D and Levine AJ. (1998) WISP genes are members of the connective tissue growth factor family that are up-regulated in wnt-1-transformed cells and aberrantly expressed in human colon tumors. *Proc. Natl. Acad. Sci. USA.* **95**(25): 14717-14722.
- Yang RB, Mark MR, Gray A, Huang A, Xie MH, Zhang M, **Goddard A**, Wood WI, Gurney AL and Godowski PJ. (1998) Toll-like receptor-2 mediates lipopolysaccharide-induced cellular signalling. *Nature* **395**(6699): 284-288.
- Merchant AM, Zhu Z, Yuan JQ, **Goddard A**, Adams CW, Presta LG and Carter P. (1998) An efficient route to human bispecific IgG. *Nature Biotechnology* **16**(7): 677-681.
- Marsters SA, Sheridan JP, Pitti RM, Brush J, **Goddard A** and Ashkenazi A. (1998) Identification of a ligand for the death-domain-containing receptor Apo3. *Current Biology* **8**(9): 525-528.
- Xie J, Murone M, Luoh SM, Ryan A, Gu Q, Zhang C, Bonifas JM, Lam CW, Hynes M, **Goddard A**, Rosenthal A, Epstein EH Jr. and de Sauvage FJ. (1998) Activating Smoothed mutations in sporadic basal-cell carcinoma. *Nature.* **391**(6662): 90-92.
- Marsters SA, Sheridan JP, Pitti RM, Huang A, Skubatch M, Baldwin D, Yuan J, Gurney A, **Goddard AD**, Godowski P and Ashkenazi A. (1997) A novel receptor for Apo2L/TRAIL contains a truncated death domain. *Current Biology.* **7**(12): 1003-1006.
- Hynes M, Stone DM, Dowd M, Pitts-Meek S, **Goddard A**, Gurney A and Rosenthal A. (1997) Control of cell pattern in the neural tube by the zinc finger transcription factor *Gli-1*. *Neuron* **19**: 15-26.
- Sheridan JP, Marsters SA, Pitti RM, Gurney A., Skubatch M, Baldwin D, Ramakrishnan L, Gray CL, Baker K, Wood WI, **Goddard AD**, Godowski P, and Ashkenazi A. (1997) Control of TRAIL-Induced Apoptosis by a Family of Signaling and Decoy Receptors. *Science* **277** (5327): 818-821.



**Goddard AD**, Dowd P, Chernausek S, Geffner M, Gertner J, Hintz R, Hopwood N, Kaplan S, Plotnick L, Rogol A, Rosenfield R, Saenger P, Mauras N, Hershkopf R, Angulo M and Attie, K. (1997) Partial growth hormone insensitivity: The role of growth hormone receptor mutations in idiopathic short stature. *J. Pediatr.* **131**: S51-55.

Klein RD, Sherman D, Ho WH, Stone D, Bennett GL, Moffat B, Vandlen R, Simmons L, Gu Q, Hongo JA, Devaux B, Poulsen K, Armanini M, Nozaki C, Asai N, **Goddard A**, Phillips H, Henderson CE, Takahashi M and Rosenthal A. (1997) A GPI-linked protein that interacts with Ret to form a candidate neurturin receptor. *Nature*. **387**(6634): 717-21.

Stone DM, Hynes M, Armanini M, Swanson TA, Gu Q, Johnson RL, Scott MP, Pennica D, **Goddard A**, Phillips H, Noll M, Hooper JE, de Sauvage F and Rosenthal A. (1996) The tumour-suppressor gene patched encodes a candidate receptor for Sonic hedgehog. *Nature* **384**(6605): 129-34.

Marsters SA, Sheridan JP, Donahue CJ, Pitti RM, Gray CL, **Goddard AD**, Bauer KD and Ashkenazi A. (1996) Apo-3, a new member of the tumor necrosis factor receptor family, contains a death domain and activates apoptosis and NF-kappa  $\beta$ . *Current Biology* **6**(12): 1669-76.

Rothe M, Xiong J, Shu HB, Williamson K, **Goddard A** and Goeddel DV. (1996) I-TRAF is a novel TRAF-interacting protein that regulates TRAF-mediated signal transduction. *Proc. Natl. Acad. Sci. USA* **93**: 8241-8246.

Yang M, Luoh SM, **Goddard A**, Reilly D, Henzel W and Bass S. (1996) The bglX gene located at 47.8 min on the Escherichia coli chromosome encodes a periplasmic beta-glucosidase. *Microbiology* **142**: 1659-65.

**Goddard AD** and Black DM. (1996) Familial Cancer in Molecular Endocrinology of Cancer. Waxman, J. Ed. Cambridge University Press, Cambridge UK, pp.187-215.

Treanor JJS, Goodman L, de Sauvage F, Stone DM, Poulson KT, Beck CD, Gray C, Armanini MP, Pollocks RA, Hefti F, Phillips HS, **Goddard A**, Moore MW, Buj-Bello A, Davis AM, Asai N, Takahashi M, Vandlen R, Henderson CE and Rosenthal A. (1996) Characterization of a receptor for GDNF. *Nature* **382**: 80-83.

Klein RD, Gu Q, **Goddard A** and Rosenthal A. (1996) Selection for genes encoding secreted proteins and receptors. *Proc. Natl. Acad. Sci. USA* **93**: 7108-7113.

Winslow JW, Moran P, Valverde J, Shih A, Yuan JQ, Wong SC, Tsai SP, **Goddard A**, Henzel WJ, Hefti F and Caras I. (1995) Cloning of AL-1, a ligand for an Eph-related tyrosine kinase receptor involved in axon bundle formation. *Neuron* **14**: 973-981.

Bennett BD, Zeigler FC, Gu Q, Fendly B, **Goddard AD**, Gillett N and Matthews W. (1995) Molecular cloning of a ligand for the EPH-related receptor protein-tyrosine kinase Htk. *Proc. Natl. Acad. Sci. USA* **92**: 1866-1870.

Huang X, Yuang J, **Goddard A**, Foulis A, James RF, Lernmark A, Pujol-Borrell R, Rabinovitch A, Somoza N and Stewart TA. (1995) Interferon expression in the pancreases of patients with type I diabetes. *Diabetes* **44**: 658-664.

**Goddard AD**, Yuan JQ, Fairbairn L, Dexter M, Borrow J, Kozak C and Solomon E. (1995) Cloning of the murine homolog of the leukemia-associated PML gene. *Mammalian Genome* **6**: 732-737.

**Goddard AD**, Covello R, Luoh SM, Clackson T, Attie KM, Gesundheit N, Rundle AC, Wells JA, Carlsson LMTI and The Growth Hormone Insensitivity Study Group. (1995) Mutations of the growth hormone receptor in children with idiopathic short stature. *N. Engl. J. Med.* **333**: 1093-1098.

Kuo SS, Moran P, Gripp J, Armanini M, Phillips HS, **Goddard A** and Caras IW. (1994) Identification and characterization of Batk, a predominantly brain-specific non-receptor protein tyrosine kinase related to Csk. *J. Neurosci. Res.* **38**: 705-715.

Mark MR, Scadden DT, Wang Z, Gu Q, **Goddard A** and Godowski PJ. (1994) Rse, a novel receptor-type tyrosine kinase with homology to Axl/Ufo, is expressed at high levels in the brain. *Journal of Biological Chemistry* **269**: 10720-10728.

Borrow J, Shipley J, Howe K, Kiely F, **Goddard A**, Sheer D, Srivastava A, Antony AC, Fioretos T, Mitelman F and Solomon E. (1994) Molecular analysis of simple variant translocations in acute promyelocytic leukemia. *Genes Chromosomes Cancer* **9**: 234-243.

**Goddard AD** and Solomon E. (1993) Genetics of Cancer. *Adv. Hum. Genet.* **21**: 321-376.

Borrow J, **Goddard AD**, Gibbons B, Katz F, Swirsky D, Fioretos T, Dube I, Winfield DA, Kingston J, Hagemeijer A, Rees JKH, Lister AT and Solomon E. (1992) Diagnosis of acute promyelocytic leukemia by RT-PCR: Detection of *PML-RARA* and *RARA-PML* fusion transcripts. *Br. J. Haematol.* **82**: 529-540.

**Goddard AD**, Borrow J and Solomon E. (1992) A previously uncharacterized gene, PML, is fused to the retinoic acid receptor alpha gene in acute promyelocytic leukemia. *Leukemia* **6 Suppl 3**: 117S-119S.

Zhu X, Dunn JM, **Goddard AD**, Squire JA, Becker A, Phillips RA and Gallie BL. (1992) Mechanisms of loss of heterozygosity in retinoblastoma. *Cytogenet. Cell. Genet.* **59**: 248-252.

Foulkes W, **Goddard A**. and Patel K. (1991) Retinoblastoma linked with Seascale [letter]. *British Med. J.* **302**: 409.

**Goddard AD**, Borrow J, Freemont PS and Solomon E. (1991) Characterization of a novel zinc finger gene disrupted by the t(15;17) in acute promyelocytic leukemia. *Science* **254**: 1371-1374.

Solomon E, Borrow J and **Goddard AD**. (1991) Chromosomal aberrations in cancer. *Science* **254**: 1153-1160.

Pajunen L, Jones TA, **Goddard A**, Sheer D, Solomon E, Pihlajaniemi T and Kivirikko KI. (1991) Regional assignment of the human gene coding for a multifunctional peptide (P4HB) acting as the  $\beta$ -subunit of prolyl-4-hydroxylase and the enzyme protein disulfide isomerase to 17q25. *Cytogenet. Cell. Genet.* **56**: 165-168.

Borrow J, Black DM, **Goddard AD**, Yagle MK, Frischauf A.-M and Solomon E. (1991) Construction and regional localization of a *NotI* linking library from human chromosome 17q. *Genomics* **10**: 477-480.

Borrow J, **Goddard AD**, Sheer D and Solomon E. (1990) Molecular analysis of acute promyelocytic leukemia breakpoint cluster region on chromosome 17. *Science* **249**: 1577-1580.

Myers JC, Jones TA, Pohjolainen E-R, Kadri AS, **Goddard AD**, Sheer D, Solomon E and Pihlajaniemi T. (1990) Molecular cloning of 5(IV) collagen and assignment of the gene to the region of the X-chromosome containing the Alport Syndrome locus. *Am. J. Hum. Genet.* **46**: 1024-1033.

Gallie BL, Squire JA, **Goddard A**, Dunn JM, Canton M, Hinton D, Zhu X and Phillips RA. (1990) Mechanisms of oncogenesis in retinoblastoma. *Lab. Invest.* **62**: 394-408.

**Goddard AD**, Phillips RA, Greger V, Passarge E, Hopping W, Gallie BL and Horsthemke B. (1990) Use of the RB1 cDNA as a diagnostic probe in retinoblastoma families. *Clinical Genetics* **37**: 117-126.

Zhu XP, Dunn JM, Phillips RA, **Goddard AD**, Paton KE, Becker A and Gallie BL. (1989) Germline, but not somatic, mutations of the RB1 gene preferentially involve the paternal allele. *Nature* **340**: 312-314.

Gallie BL, Dunn JM, **Goddard A**, Becker A and Phillips RA. (1988) Identification of mutations in the putative retinoblastoma gene. In Molecular Biology of The Eye: Genes, Vision and Ocular Disease. UCLA Symposia on Molecular and Cellular Biology, New Series, Volume 88. J. Piatigorsky, T. Shinohara and P.S. Zelenka, Eds. Alan R. Liss, Inc., New York, 1988, pp. 427-436.

**Goddard AD**, Balakier H, Canton M, Dunn J, Squire J, Reyes E, Becker A, Phillips RA and Gallie BL. (1988) Infrequent genomic rearrangement and normal expression of the putative RB1 gene in retinoblastoma tumors. *Mol. Cell. Biol.* **8**: 2082-2088.

Squire J, Dunn J, **Goddard A**, Hoffman T, Musarella M, Willard HF, Becker AJ, Gallie BL and Phillips RA. (1986) Cloning of the esterase D gene: A polymorphic gene probe closely linked to the retinoblastoma locus on chromosome 13. *Proc. Natl. Acad. Sci. USA* **83**: 6573-6577.

Squire J, **Goddard AD**, Canton M, Becker A, Phillips RA and Gallie BL (1986) Tumour induction by the retinoblastoma mutation is independent of N-myc expression. *Nature* **322**: 555-557.

**Goddard AD**, Heddle JA, Gallie BL and Phillips RA. (1985) Radiation sensitivity of fibroblasts of bilateral retinoblastoma patients as determined by micronucleus induction *in vitro*. *Mutation Research* **152**: 31-38.

## RESEARCH

## SIMULTANEOUS AMPLIFICATION AND DETECTION OF SPECIFIC DNA SEQUENCES

Russell Higuchi\*, Gavin Dollinger<sup>1</sup>, P. Sean Walsh and Robert GriffithRoche Molecular Systems, Inc., 1400 53rd St., Emeryville, CA 94608. <sup>1</sup>Chiron Corporation, 1400 53rd St., Emeryville, CA 94608. \*Corresponding author.

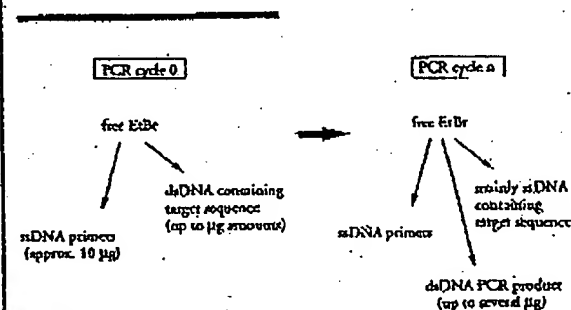
We have enhanced the polymerase chain reaction (PCR) such that specific DNA sequences can be detected without opening the reaction tube. This enhancement requires the addition of ethidium bromide (EtBr) to a PCR. Since the fluorescence of EtBr increases in the presence of double-stranded (ds) DNA an increase in fluorescence in such a PCR indicates a positive amplification, which can be easily monitored externally. In fact, amplification can be continuously monitored in order to follow its progress. The ability to simultaneously amplify specific DNA sequences and detect the product of the amplification both simplifies and improves PCR and may facilitate its automation and more widespread use in the clinic or in other situations requiring high sample throughput.

Although the potential benefits of PCR<sup>1</sup> to clinical diagnostics are well known<sup>2,3</sup>, it is still not widely used in this setting, even though it is four years since thermostable DNA polymerases<sup>4</sup> made PCR practical. Some of the reasons for its slow acceptance are high cost, lack of automation of pre- and post-PCR processing steps, and false positive results from carryover-contamination. The first two points are related in that labor is the largest contributor to cost at the present state of PCR development. Most current assays require some form of "downstream" processing once thermocycling is done in order to determine whether the target DNA sequence was present and has amplified. These include DNA hybridization<sup>5,6</sup>, gel electrophoresis with or without use of restriction digestion<sup>7,8</sup>, HPLC<sup>9</sup>, or capillary electrophoresis<sup>10</sup>. These methods are labor-intensive, have low throughput, and are difficult to automate. The third point is also closely related to downstream processing. The handling of the PCR product in these downstream processes increases the chances that amplified DNA will spread through the typing lab, resulting in a risk of

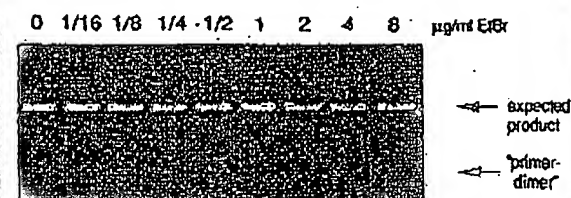
"carryover" false positives in subsequent testing<sup>11</sup>.

These downstream processing steps would be eliminated if specific amplification and detection of amplified DNA took place simultaneously within an unopened reaction vessel. Assays in which such different processes take place without the need to separate reaction components have been termed "homogeneous". No truly homogeneous PCR assay has been demonstrated to date, although progress towards this end has been reported. Chehab, et al.<sup>12</sup>, developed a PCR product detection scheme using fluorescent primers that resulted in a fluorescent PCR product. Allele-specific primers, each with different fluorescent tags, were used to indicate the genotype of the DNA. However, the unincorporated primers must still be removed in a downstream process in order to visualize the result. Recently, Holland, et al.<sup>13</sup>, developed an assay in which the endogenous 5' exonuclease assay of *Taq* DNA polymerase was exploited to cleave a labeled oligonucleotide probe. The probe would only cleave if PCR amplification had produced its complementary sequence. In order to detect the cleavage products, however, a subsequent process is again needed.

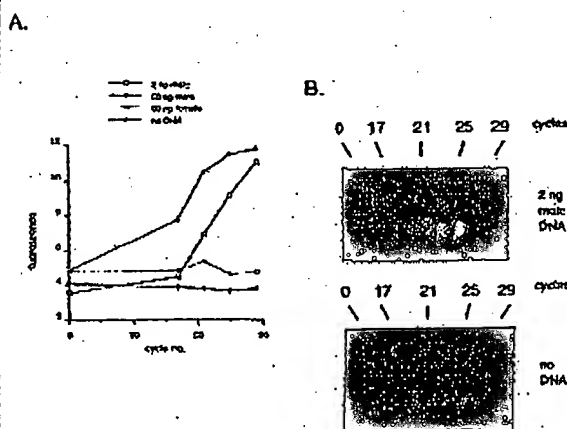
We have developed a truly homogeneous assay for PCR and PCR product detection based upon the greatly increased fluorescence that ethidium bromide and other DNA binding dyes exhibit when they are bound to ds-DNA<sup>14-16</sup>. As outlined in Figure 1, a prototypic PCR



**FIGURE 1** Principle of simultaneous amplification and detection of PCR product. The components of a PCR containing EtBr that are fluorescent are listed—EtBr itself, EtBr bound to either ssDNA or dsDNA. There is a large fluorescence enhancement when EtBr is bound to DNA and binding is greatly enhanced when DNA is double-stranded. After sufficient (n) cycles of PCR, the net increase in dsDNA results in additional EtBr binding, and a net increase in total fluorescence.



**FIGURE 2** Gel electrophoresis of PCR amplification products of the human nuclear gene, HLA DQα, made in the presence of increasing amounts of EtBr (up to 8 μg/ml). The presence of EtBr has no obvious effect on the yield or specificity of amplification.



**FIGURE 3** (A) Fluorescence measurements from PCR reactions that contain 0.5 μg/ml EtBr and that are specific for Y-chromosome repeat sequences. Five replicate PCRs were begun containing each of the DNAs specified. At each indicated cycle, one of the five replicate PCRs for each DNA was removed from thermocycling and its fluorescence measured. Units of fluorescence are arbitrary. (B) UV photography of PCR tubes (0.5 ml Eppendorf-style, polypropylene micro-centrifuge tubes) containing reactions, those starting from 2 ng male DNA and control reactions without any DNA, from (A).

begins with primers that are single-stranded DNA (ss-DNA), dNTPs, and DNA polymerase. An amount of dsDNA containing the target sequence (target DNA) is also typically present. This amount can vary, depending on the application, from single-cell amounts of DNA<sup>17</sup> to micrograms per PCR<sup>18</sup>. If EtBr is present, the reagents that will fluoresce, in order of increasing fluorescence, are free EtBr itself, and EtBr bound to the single-stranded DNA primers and to the double-stranded target DNA (by its intercalation between the stacked bases of the DNA double-helix). After the first denaturation cycle, target DNA will be largely single-stranded. After a PCR is completed, the most significant change is the increase in the amount of dsDNA (the PCR product itself) of up to several micrograms. Formerly free EtBr is bound to the additional dsDNA, resulting in an increase in fluorescence. There is also some decrease in the amount of ssDNA primer, but because the binding of EtBr to ssDNA is much less than to dsDNA, the effect of this change on the total fluorescence of the sample is small. The fluorescence increase can be measured by directing excitation illumination through the walls of the amplification vessel

before and after, or even continuously during, thermocycling.

## RESULTS

**PCR in the presence of EtBr.** In order to assess the effect of EtBr in PCR, amplifications of the human HLA DQα gene<sup>19</sup> were performed with the dye present at concentrations from 0.06 to 8.0 μg/ml (a typical concentration of EtBr used in staining of nucleic acids following gel electrophoresis is 0.5 μg/ml). As shown in Figure 2, gel electrophoresis revealed little or no difference in the yield or quality of the amplification product whether EtBr was absent or present at any of these concentrations, indicating that EtBr does not inhibit PCR.

**Detection of human Y-chromosome specific sequences.** Sequence-specific, fluorescence enhancement of EtBr as a result of PCR was demonstrated in a series of amplifications containing 0.5 μg/ml EtBr and primers specific to repeat DNA sequences found on the human Y-chromosome<sup>20</sup>. These PCRs initially contained either 60 ng male, 60 ng female, 2 ng male human or no DNA. Five replicate PCRs were begun for each DNA. After 0, 17, 21, 24 and 29 cycles of thermocycling, a PCR for each DNA was removed from the thermocycler, and its fluorescence measured in a spectrofluorometer and plotted vs. amplification cycle number (Fig. 3A). The shape of this curve reflects the fact that by the time an increase in fluorescence can be detected, the increase in DNA is becoming linear and not exponential with cycle number. As shown, the fluorescence increased about three-fold over the background fluorescence for the PCRs containing human male DNA, but did not significantly increase for negative control PCRs, which contained either no DNA or human female DNA. The more male DNA present to begin with—60 ng versus 2 ng—the fewer cycles were needed to give a detectable increase in fluorescence. Gel electrophoresis on the products of these amplifications showed that DNA fragments of the expected size were made in the male DNA containing reactions and that little DNA synthesis took place in the control samples.

In addition, the increase in fluorescence was visualized by simply laying the completed, unopened PCRs on a UV transilluminator and photographing them through a red filter. This is shown in figure 3B for the reactions that began with 2 ng male DNA and those with no DNA.

**Detection of specific alleles of the human β-globin gene.** In order to demonstrate that this approach has adequate specificity to allow genetic screening, a detection of the sickle-cell anemia mutation was performed. Figure 4 shows the fluorescence from completed amplifications containing EtBr (0.5 μg/ml) as detected by photography of the reaction tubes on a UV transilluminator. These reactions were performed using primers specific for either the wild-type or sickle-cell mutation of the human β-globin gene<sup>21</sup>. The specificity for each allele is imparted by placing the sickle-mutation site at the terminal 3' nucleotide of one primer. By using an appropriate primer annealing temperature, primer extension—and thus amplification—can take place only if the 3' nucleotide of the primer is complementary to the β-globin allele present.<sup>21,22</sup>

Each pair of amplifications shown in Figure 4 consists of a reaction with either the wild-type allele specific (left tube) or sickle-allele specific (right tube) primers. Three different DNAs were typed: DNA from a homozygous, wild-type β-globin individual (AA); from a heterozygous sickle β-globin individual (AS); and from a homozygous sickle β-globin individual (SS). Each DNA (50 ng genomic DNA to start each PCR) was analyzed in triplicate (3 pairs

of reactions each). The DNA type was reflected in the relative fluorescence intensities in each pair of completed amplifications. There was a significant increase in fluorescence only where a  $\beta$ -globin allele DNA matched the primer set. When measured on a spectrofluorometer (data not shown), this fluorescence was about three times that present in a PCR where both  $\beta$ -globin alleles were mismatched to the primer set. Gel electrophoresis (not shown) established that this increase in fluorescence was due to the synthesis of nearly a microgram of a DNA fragment of the expected size for  $\beta$ -globin. There was little synthesis of dsDNA in reactions in which the allele-specific primer was mismatched to both alleles.

**Continuous monitoring of a PCR.** Using a fiber optic device, it is possible to direct excitation illumination from a spectrofluorometer to a PCR undergoing thermocycling and to return its fluorescence to the spectrofluorometer. The fluorescence readout of such an arrangement, directed at an EtBr-containing amplification of Y-chromosome specific sequences from 25 ng of human male DNA, is shown in Figure 5. The readout from a control PCR with no target DNA is also shown. Thirty cycles of PCR were monitored for each.

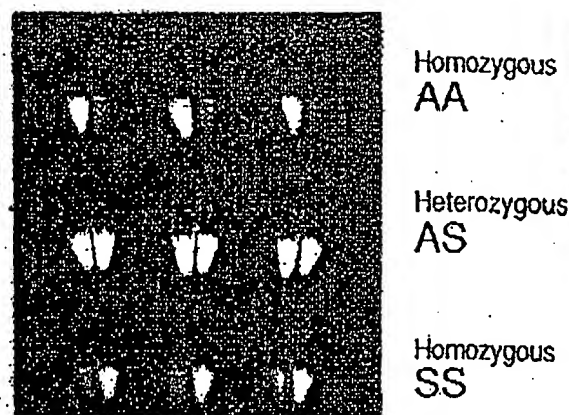
The fluorescence trace as a function of time clearly shows the effect of the thermocycling. Fluorescence intensity rises and falls inversely with temperature. The fluorescence intensity is minimum at the denaturation temperature (94°C) and maximum at the annealing/extension temperature (50°C). In the negative-control PCR, these fluorescence maxima and minima do not change significantly over the thirty thermocycles, indicating that there is little dsDNA synthesis without the appropriate target DNA, and there is little if any bleaching of EtBr during the continuous illumination of the sample.

In the PCR containing male DNA, the fluorescence maxima at the annealing/extension temperature begin to increase at about 4000 seconds of thermocycling, and continue to increase with time, indicating that dsDNA is being produced at a detectable level. Note that the fluorescence minima at the denaturation temperature do not significantly increase, presumably because at this temperature there is no dsDNA for EtBr to bind. Thus the course of the amplification is followed by tracking the fluorescence increase at the annealing temperature. Analysis of the products of these two amplifications by gel electrophoresis showed a DNA fragment of the expected size for the male DNA containing sample and no detectable DNA synthesis for the control sample.

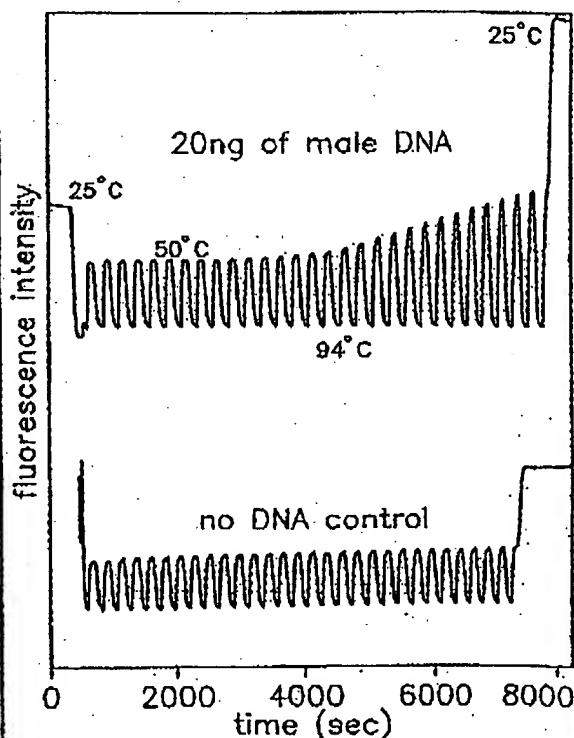
## DISCUSSION

Downstream processes such as hybridization to a sequence-specific probe can enhance the specificity of DNA detection by PCR. The elimination of these processes means that the specificity of this homogeneous assay depends solely on that of PCR. In the case of sickle-cell disease, we have shown that PCR alone has sufficient DNA sequence specificity to permit genetic screening. Using appropriate amplification conditions, there is little non-specific production of dsDNA in the absence of the appropriate target allele.

The specificity required to detect pathogens can be more or less than that required to do genetic screening, depending on the number of pathogens in the sample and the amount of other DNA that must be taken with the sample. A difficult target is HIV, which requires detection of a viral genome that can be at the level of a few copies per thousands of host cells<sup>6</sup>. Compared with genetic screening, which is performed on cells containing at least one copy of the target sequence, HIV detection requires both more specificity and the input of more total



**FIGURE 4** UV photograph of PCR tubes containing amplifications using EtBr that are specific to wild-type (A) or sickle (S) alleles of the human  $\beta$ -globin gene. The left of each pair of tubes contains allele-specific primers to the wild-type alleles, the right tube primers to the sickle allele. The photograph was taken after 30 cycles of PCR, and the input DNAs and the alleles they contain are indicated. Fifty ng of DNA was used to begin PCR. Typing was done in triplicate (3 pairs of PCRs) for each input DNA.



**FIGURE 5** Continuous, real-time monitoring of a PCR. A fiber optic was used to carry excitation light to a PCR in progress and also emitted light back to a fluorometer (see Experimental Protocol). Amplification using human male-DNA specific primers in a PCR starting with 20 ng of human male DNA (top), or in a control PCR without DNA (bottom), were monitored. Thirty cycles of PCR were followed for each. The temperature cycled between 94°C (denaturation) and 50°C (annealing and extension). Note in the male DNA PCR, the cycle (time) dependent increase in fluorescence at the annealing/extension temperature.



DNA—up to microgram amounts—in order to have sufficient numbers of target sequences. This large amount of starting DNA in an amplification significantly increases the background fluorescence over which any additional fluorescence produced by PCR must be detected. An additional complication that occurs with targets in low copy-number is the formation of the "primer-dimer" artifact. This is the result of the extension of one primer using the other primer as a template. Although this occurs infrequently, once it occurs the extension product is a substrate for PCR amplification, and can compete with true PCR targets if those targets are rare. The primer-dimer product is of course dsDNA and thus is a potential source of false signal in this homogeneous assay.

To increase PCR specificity and reduce the effect of primer-dimer amplification, we are investigating a number of approaches, including the use of nested-primer amplifications that take place in a single tube<sup>9</sup>, and the "hot-start", in which nonspecific amplification is reduced by raising the temperature of the reaction before DNA synthesis begins<sup>23</sup>. Preliminary results using these approaches suggest that primer-dimer is effectively reduced and it is possible to detect the increase in EtBr fluorescence in a PCR instigated by a single HIV genome in a background of  $10^5$  cells. With larger numbers of cells, the background fluorescence contributed by genomic DNA becomes problematic. To reduce this background, it may be possible to use sequence-specific DNA-binding dyes that can be made to preferentially bind PCR product over genomic DNA by incorporating the dye-binding DNA sequence into the PCR product through a 5' "add-on" to the oligonucleotide primer<sup>24</sup>.

We have shown that the detection of fluorescence generated by an EtBr-containing PCR is straightforward, both once PCR is completed and continuously during thermocycling. The ease with which automation of specific DNA detection can be accomplished is the most promising aspect of this assay. The fluorescence analysis of completed PCRs is already possible with existing instrumentation in 96-well format<sup>25</sup>. In this format, the fluorescence in each PCR can be quantitated before, after, and even at selected points during thermocycling by moving the rack of PCRs to a 96-microwell plate fluorescence reader<sup>26</sup>.

The instrumentation necessary to continuously monitor multiple PCRs simultaneously is also simple in principle. A direct extension of the apparatus used here is to have multiple fiberoptics transmit the excitation light and fluorescent emissions to and from multiple PCRs. The ability to monitor multiple PCRs continuously may allow quantitation of target DNA copy number. Figure 3 shows that the larger the amount of starting target DNA, the sooner during PCR a fluorescence increase is detected. Preliminary experiments (Higuchi and Dollinger, manuscript in preparation) with continuous monitoring have shown a sensitivity to two-fold differences in initial target DNA concentration.

Conversely, if the number of target molecules is known—as it can be in genetic screening—continuous monitoring may provide a means of detecting false positive and false negative results. With a known number of target molecules, a true positive would exhibit detectable fluorescence by a predictable number of cycles of PCR. Increases in fluorescence detected before or after that cycle would indicate potential artifacts. False negative results due to, for example, inhibition of DNA polymerase, may be detected by including within each PCR an inefficiently amplifying marker. This marker results in a fluorescence increase only after a large number of cycles—many more than are necessary to detect a true

positive. If a sample fails to have a fluorescence increase after this many cycles, inhibition may be suspected. Since, in this assay, conclusions are drawn based on the presence or absence of fluorescence signal alone, such controls may be important. In any event, before any test based on this principle is ready for the clinic, an assessment of its false positive/false negative rates will need to be obtained using a large number of known samples.

In summary, the inclusion in PCR of dyes whose fluorescence is enhanced upon binding dsDNA makes it possible to detect specific DNA amplification from outside the PCR tube. In the future, instruments based upon this principle may facilitate the more widespread use of PCR in applications that demand the high throughput of samples.

#### EXPERIMENTAL PROTOCOL

**Human HLA-DQ $\alpha$  gene amplifications containing EtBr.** PCRs were set up in 100  $\mu$ l volumes containing 10 mM Tris-HCl, pH 8.3; 50 mM KCl; 4 mM MgCl<sub>2</sub>; 2.5 units of Taq DNA polymerase (Perkin-Elmer Cetus, Norwalk, CT); 20 pmole each of human HLA-DQ $\alpha$  gene specific oligonucleotide primers (H26 and H27<sup>19</sup> and approximately  $10^5$  copies of DQ $\alpha$  PCR product diluted from a previous reaction. Ethidium bromide (EtBr; Sigma) was used at the concentrations indicated in Figure 2. Thermocycling proceeded for 20 cycles in a model 480 thermocycler (Perkin-Elmer Cetus, Norwalk, CT) using a "step-cycle" program of 94°C for 1 min., denaturation and 60°C for 30 sec. annealing and 72°C for 30 sec. extension.

**Y-chromosome specific PCR.** PCRs (100  $\mu$ l total reaction volume) containing 0.5  $\mu$ g/ml EtBr were prepared as described for HLA-DQ $\alpha$ , except with different primers and target DNAs. These PCRs contained 15 pmole each male DNA-specific primers Y1.1 and Y1.2<sup>20</sup>, and either 60 ng male, 60 ng female, 2 ng male, or no human DNA. Thermocycling was 94°C for 1 min. and 60°C for 1 min using a "step-cycle" program. The number of cycles for a sample were as indicated in Figure 3. Fluorescence measurement is described below.

**Allele-specific, human  $\beta$ -globin gene PCR.** Amplifications of 100  $\mu$ l volume using 0.5  $\mu$ g/ml of EtBr were prepared as described for HLA-DQ $\alpha$  above except with different primers and target DNAs. These PCRs contained either primer pair HGP2/HB14A (wild-type globin specific primers) or HGP2/HB14S (sickle-globin specific primers) at 10 pmole each primer per PCR. These primers were developed by Wu et al.<sup>21</sup>. Three different target DNAs were used in separate amplifications—50 ng each of human DNA that was homozygous for the sickle trait (SS), DNA that was heterozygous for the sickle trait (AS), or DNA that was homozygous for the w.t. globin (AA). Thermocycling was for 30 cycles at 94°C for 1 min. and 55°C for 1 min. using a "step-cycle" program. An annealing temperature of 55°C had been shown by Wu et al.<sup>21</sup> to provide allele-specific amplification. Completed PCRs were photographed through a red filter (Wratten 23A) after placing the reaction tubes atop a model TM-36 transilluminator (UV-products San Gabriel, CA).

**Fluorescence measurement.** Fluorescence measurements were made on PCRs containing EtBr in a Fluorolog-2 fluorometer (SPEX, Edison, NJ). Excitation was at the 500 nm band with about 2 nm bandwidth with a GG 435 nm cut-off filter (Melles Griest, Inc., Irvine, CA) to exclude second-order light. Emitted light was detected at 570 nm with a bandwidth of about 7 nm. An OG 530 nm cut-off filter was used to remove the excitation light.

**Continuous fluorescence monitoring of PCR.** Continuous monitoring of a PCR in progress was accomplished using the spectrofluorometer and settings described above as well as a fiberoptic accessory (SPEX cat. no. 1950) to both send excitation light to, and receive emitted light from, a PCR placed in a well of a model 480 thermocycler (Perkin-Elmer Cetus). The probe end of the fiberoptic cable was attached with "5 minute-epoxy" to the open top of a PCR tube (a 0.5 ml polypropylene centrifuge tube with its cap removed) effectively sealing it. The exposed top of the PCR tube and the end of the fiberoptic cable were shielded from room light and the room lights were kept dimmed during each run. The monitored PCR was an amplification of Y-chromosome-specific repeat sequences as described above, except using an annealing/extension temperature of 50°C. The reaction was covered with mineral oil (2 drops) to prevent evaporation. Thermocycling and fluorescence measurement were started simultaneously. A time-base scan with a 10 second integration time

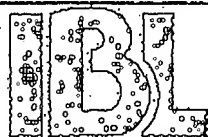
was used and the emission signal was ratioed to the excitation signal to control for changes in light-source intensity. Data were collected using the dm3000f, version 2.5 (SPEX) data system.

#### Acknowledgments

We thank Bob Jones for help with the spectrofluorometric measurements and Heatherbell Fong for editing this manuscript.

#### References

- Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H. 1986. Specific enzymatic amplification of DNA *in vitro*: The polymerase chain reaction. *CSHSQB* 51:263-273.
- White, T. J., Arnheim, N. and Erlich, H. A. 1989. The polymerase chain reaction. *Trends Genet.* 5:185-189.
- Erlich, H. A., Gelfand, D. and Smirsky, J. J. 1991. Recent advances in the polymerase chain reaction. *Science* 252:1643-1651.
- Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239:487-491.
- Saiki, R. K., Walsh, P. S., Levenson, C. H. and Erlich, H. A. 1989. Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. *Proc. Natl. Acad. Sci. USA* 86:6230-6234.
- Kwok, S. Y., Mack, D. H., Mullis, K. B., Poiesz, B. J., Ehrlich, G. D., Blair, D. and Friedman-Kien, A. S. 1987. Identification of human immunodeficiency virus sequences by using *in vitro* enzymatic amplification and oligomer cleavage detection. *J. Virol.* 61:1690-1694.
- Chehab, F. F., Doherty, M., Cai, S. P., Kan, Y. W., Cooper, S. and Rubin, E. M. 1987. Detection of sickle cell anemia and thalassemia. *Nature* 329:203-204.
- Horn, G. T., Richards, B. and Klingner, R. W. 1989. Amplification of a highly polymorphic VNTR segment by the polymerase chain reaction. *Nuc. Acids Res.* 16:2140.
- Katz, E. and Dong, M. W. 1990. Rapid analysis and purification of polymerase chain reaction products by high-performance liquid chromatography. *Biotechniques* 8:546-555.
- Helger, D. N., Cohen, A. S. and Rager, B. L. 1990. Separation of DNA restriction fragments by high performance capillary electrophoresis with low and zero crosslinked polyacrylamide using continuous and pulsed electric fields. *J. Chromatogr.* 516:33-48.
- Kwok, S. Y. and Higuchi, R. G. 1989. Avoiding false positives with PCR. *Nature* 339:237-238.
- Chehab, F. F. and Kan, Y. W. 1989. Detection of specific DNA sequences by fluorescence amplification: a color complementation assay. *Proc. Natl. Acad. Sci. USA* 86:9178-9182.
- Holland, P. M., Abramson, R. D., Watson, R. and Gelfand, D. H. 1991. Detection of specific polymerase chain reaction products by utilizing the 5' to 3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. Natl. Acad. Sci. USA* 88:7276-7280.
- Markovits, J., Roques, B. P. and Le Pecq, J. B. 1979. Ethidium dimer: a new reagent for the fluorimetric determination of nucleic acids. *Anal. Biochem.* 94:259-264.
- Kapuscinski, J. and Sact, W. 1979. Interactions of 4',6-diamidino-2-phenylindole with synthetic polynucleotides. *Nuc. Acids Res.* 6:5519-5534.
- Searle, M. S. and Embrey, K. J. 1990. Sequence-specific interaction of Hoechst 33258 with the minor groove of an adenine-tract DNA duplex studied in solution by <sup>1</sup>H NMR spectroscopy. *Nuc. Acids Res.* 18:3753-3762.
- Li, H. H., Gyllenstein, U. B., Cui, X. F., Saiki, R. K., Erlich, H. A. and Arnheim, N. 1988. Amplification and analysis of DNA sequences in single human sperm and diploid cells. *Nature* 336:414-417.
- Abbott, M. A., Poiesz, B. J., Byrne, B. C., Kwok, S. Y., Smirsky, J. J. and Erlich, H. A. 1988. Enzymatic gene amplification: qualitative and quantitative methods for detecting proviral DNA amplified *in vitro*. *J. Infect. Dis.* 158:1158.
- Saiki, R. K., Bugawan, T. L., Horn, G. T., Mullis, K. B. and Erlich, H. A. 1986. Analysis of enzymatically amplified  $\beta$ -globin and HLA-DQA DNA with allele-specific oligonucleotide probes. *Nature* 324:163-166.
- Eogan, S. C., Doherty, M. and Giocchio, J. 1987. An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences. *N. Engl. J. Med.* 317:985-990.
- Wu, D. Y., Ugozaka, L., Pal, B. R. and Wallace, R. B. 1989. Allele-specific enzymatic amplification of  $\beta$ -globin genomic DNA for diagnosis of sickle cell anemia. *Proc. Natl. Acad. Sci. USA* 86:2757-2760.
- Kwok, S., Kellogg, D. E., McKinney, N., Spasic, D., Goda, L., Levenson, C. and Smirsky, J. J. 1990. Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. *Nuc. Acids Res.* 18:999-1005.
- Chou, Q., Russell, M., Birch, D., Raymond, J. and Bloch, W. 1992. Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications: Submitted.
- Higuchi, R. 1989. Using PCR to engineer DNA. p. 61-70. In: *PCR Technology*. H. A. Erlich (Ed.). Stockton Press, New York, N.Y.
- Hall, L., Atwood, J. G., DiCesare, J., Katz, E., Pontaza, E., Williams, J. F. and Wontenberg, T. 1991. A high-performance system for automation of the polymerase chain reaction. *Biotechniques* 10:102-109, 106-112.
- Tamura, N. and Kalwa, L. 1989. Fluorescent ELISA screening of monoclonal antibodies to cell surface antigens. *J. Immun. Med.* 116:59-63.



IMMUNO BIOLOGICAL LABORATORIES

## SCD-14 ELISA

# Trauma, Shock and Sepsis

The CD-14 molecule is expressed on the surface of monocytes and some macrophages. Membrane-bound CD-14 is a receptor for lipopolysaccharide (LPS) complexed to LPS-Binding-Protein (LBP). The concentration of its soluble form is altered under certain pathological conditions. There is evidence for an important role of sCD-14 with polytrauma, sepsis, burnings and inflammations. During septic conditions and acute infections it seems to be a prognostic marker and is therefore of value in monitoring these patients.

IBL offers an ELISA for quantitative determination of soluble CD-14 in human serum, -plasma, cell-culture supernatants and other biological fluids.

Assay features: 12x8 determinations (microtiter strips),  
precoated with a specific monoclonal antibody,  
2x1 hour incubation,  
standard range: 3 - 96 ng/ml  
detection limit: 1 ng/ml  
CV: intra- and interassay < 8%

For more information call or fax

GESELLSCHAFT FÜR IMMUNCHEMIE UND -BIOLOGIE MBH  
OSTERSTRASSE 86 · D · 2000 HAMBURG 20 · GERMANY · TEL. +40/491 00 61-64 · FAX +40/40 11 98

BIOTECHNOLOGY VOL 10 APR 1992



# Oligonucleotides with Fluorescent Dyes at Opposite Ends Provide a Quenched Probe System Useful for Detecting PCR Product and Nucleic Acid Hybridization

Kenneth J. Livak, Susan J.A. Flood, Jeffrey Marmaro, William Giusti, and Karin Deetz

Perkin-Elmer, Applied Biosystems Division, Foster City, California 94404

The 5' nuclease PCR assay detects the accumulation of specific PCR product by hybridization and cleavage of a double-labeled fluorogenic probe during the amplification reaction. The probe is an oligonucleotide with both a reporter fluorescent dye and a quencher dye attached. An increase in reporter fluorescence intensity indicates that the probe has hybridized to the target PCR product and has been cleaved by the 5' → 3' nucleolytic activity of *Taq* DNA polymerase. In this study, probes with the quencher dye attached to an internal nucleotide were compared with probes with the quencher dye attached to the 3'-end nucleotide. In all cases, the reporter dye was attached to the 5' end. All intact probes showed quenching of the reporter fluorescence. In general, probes with the quencher dye attached to the 3'-end nucleotide exhibited a larger signal in the 5' nuclease PCR assay than the internally labeled probes. It is proposed that the larger signal is caused by increased likelihood of cleavage by *Taq* DNA polymerase when the probe is hybridized to a template strand during PCR. Probes with the quencher dye attached to the 3'-end nucleotide also exhibited an increase in reporter fluorescence intensity when hybridized to a complementary strand. Thus, oligonucleotides with reporter and quencher dyes attached at opposite ends can be used as homogeneous hybridization probes.

A homogeneous assay for detecting the accumulation of specific PCR product that uses a double-labeled fluorogenic probe was described by Lee et al.<sup>(1)</sup> The assay exploits the 5' → 3' nucleolytic activity of *Taq* DNA polymerase<sup>(2,3)</sup> and is diagramed in Figure 1. The fluorogenic probe consists of an oligonucleotide with a reporter fluorescent dye, such as a fluorescein, attached to the 5' end; and a quencher dye, such as a rhodamine, attached internally. When the fluorescein is excited by irradiation, its fluorescent emission will be quenched if the rhodamine is close enough to be excited through the process of fluorescence energy transfer (FET).<sup>(4,5)</sup> During PCR, if the probe is hybridized to a template strand, *Taq* DNA polymerase will cleave the probe because of its inherent 5' → 3' nucleolytic activity. If the cleavage occurs between the fluorescein and rhodamine dyes, it causes an increase in fluorescein fluorescence intensity because the fluorescein is no longer quenched. The increase in fluorescein fluorescence intensity indicates that the probe-specific PCR product has been generated. Thus, FET between a reporter dye and a quencher dye is critical to the performance of the probe in the 5' nuclease PCR assay.

Quenching is completely dependent on the physical proximity of the two dyes.<sup>(6)</sup> Because of this, it has been assumed that the quencher dye must be attached near the 5' end. Surprisingly, we have found that attaching a rhodamine dye at the 3' end of a probe still provides adequate quenching for the probe to perform in the 5' nuclease

PCR assay. Furthermore, cleavage of this type of probe is not required to achieve some reduction in quenching. Oligonucleotides with a reporter dye on the 5' end and a quencher dye on the 3' end exhibit a much higher reporter fluorescence when double-stranded as compared with single-stranded. This should make it possible to use this type of double-labeled probe for homogeneous detection of nucleic acid hybridization.

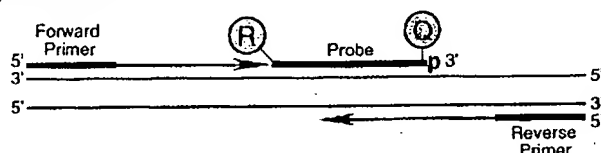
## MATERIALS AND METHODS

### Oligonucleotides

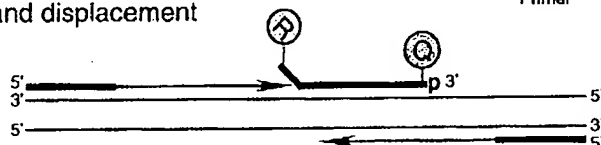
Table 1 shows the nucleotide sequence of the oligonucleotides used in this study. Linker arm nucleotide (LAN) phosphoramidite was obtained from Glen Research. The standard DNA phosphoramidites, 6-carboxyfluorescein (6-FAM) phosphoramidite, 6-carboxytetramethylrhodamine succinimidyl ester (TAMRA NHS ester), and Phosphalink for attaching a 3'-blocking phosphate, were obtained from Perkin-Elmer, Applied Biosystems Division. Oligonucleotide synthesis was performed using an ABI model 394 DNA synthesizer (Applied Biosystems). Primer and complement oligonucleotides were purified using Oligo Purification Cartridges (Applied Biosystems). Double-labeled probes were synthesized with 6-FAM-labeled phosphoramidite at the 5' end, LAN replacing one of the T's in the sequence, and Phosphalink at the 3' end. Following deprotection and ethanol precipitation, TAMRA NHS ester was coupled to the LAN-containing oligonucleotide in 250

# Research

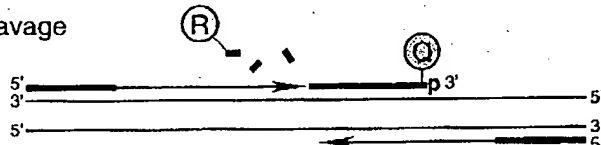
## Polymerization



## Strand displacement



## Cleavage



## Polymerization completed

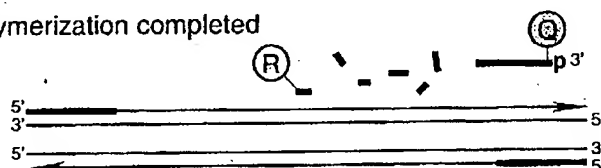


FIGURE 1 Diagram of 5' nuclease assay. Stepwise representation of the 5' → 3' nucleolytic activity of *Taq* DNA polymerase acting on a fluorogenic probe during one extension phase of PCR.

mm Na-bicarbonate buffer (pH 9.0) at room temperature. Unreacted dye was removed by passage over a PD-10 Sephadex column. Finally, the double-labeled probe was purified by preparative high-performance liquid chromatography (HPLC) using an Aquapore C<sub>8</sub> 220×4.6-mm column with 7-μm particle size. The column was developed with a 24-min linear gradient of 8–20% acetonitrile in 0.1 M TEAA (triethylamine acetate). Probes are named by designating the sequence from Table 1 and the position of the LAN-TAMRA moiety. For example, probe A1-7 has sequence A1 with LAN-TAMRA at nucleotide position 7 from the 5' end.

### PCR Systems

All PCR amplifications were performed in the Perkin-Elmer GeneAmp PCR System 9600 using 50-μl reactions that contained 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 200 μM dATP, 200 μM dCTP, 200 μM dGTP, 400 μM dUTP, 0.5 unit of AmpErase uracil N-glycosylase (Perkin-Elmer), and 1.25 unit of AmpliTaq DNA polymerase (Perkin-Elmer). A 295-bp segment from exon 3 of the human β-actin

gene (nucleotides 2141–2435 in the sequence of Nakajima-Iijima et al.<sup>(7)</sup>) was amplified using primers AFP and ARP (Table 1), which are modified slightly from those of du Breuil et al.<sup>(8)</sup> Actin amplification reactions contained 4 mM MgCl<sub>2</sub>, 20 ng of human genomic DNA, 50 nM A1 or A3 probe, and 300 nM each

primer. The thermal regimen was 50°C (2 min), 95°C (10 min), 40 cycles of 95°C (20 sec), 60°C (1 min), and hold at 72°C. A 515-bp segment was amplified from a plasmid that consists of a segment of λ DNA (nucleotides 32,220–32,747) inserted in the *Sma*I site of vector pUC119. These reactions contained 3.5 mM MgCl<sub>2</sub>, 1 ng of plasmid DNA, 50 nM P2 or P5 probe, 200 nM primer F119, and 200 nM primer R119. The thermal regimen was 50°C (2 min), 95°C (10 min), 25 cycles of 95°C (20 sec), 57°C (1 min), and hold at 72°C.

### Fluorescence Detection

For each amplification reaction, a 40-μl aliquot of a sample was transferred to an individual well of a white, 96-well microtiter plate (Perkin-Elmer). Fluorescence was measured on the Perkin-Elmer TaqMan LS-50B System, which consists of a luminescence spectrometer with plate reader assembly, a 485-nm excitation filter, and a 515-nm emission filter. Excitation was at 488 nm using a 5-nm slit width. Emission was measured at 518 nm for 6-FAM (the reporter or R value) and 582 nm for TAMRA (the quencher or Q value) using a 10-nm slit width. To determine the increase in reporter emission that is caused by cleavage of the probe during PCR, three normalizations are applied to the raw emission data. First, emission intensity of a buffer blank is subtracted for each wavelength. Second, emission intensity of the reporter is

TABLE 1 Sequences of Oligonucleotides

Name	Type	Sequence
F119	primer	ACCCACAGGAACATGATCACCACCTC
R119	primer	ATGTCGCGTTCGGCTGACGTTCTGC
P2	probe	TGCGATTACGATCGTTGCCAACCAGTp
P2C	complement	GTACTGGTTGGCAACGATCAGTAATGCGATG
P5	probe	CGGAATTGCTGGTATCTATGACAAGGATp
P5C	complement	TTTCATCCTTGTATAGATACCAAGCAATCCG
AFP	primer	TCACCCACACTGTGCCCATCTACGA
ARP	primer	CAGCGGAACCGCTCATTGCCAATGG
A1	probe	ATGCCCTCCCCATGCCATCCCTCGTp
A1C	complement	AGACGCAGGATGGCATGGGGGAGGGCATAC
A3	probe	CGCCCCTGGACTTCGAGCAAGAGATp
A3C	complement	CCATCTCTTGCTCGAAGTCCAGGGCGAC

For each oligonucleotide used in this study, the nucleic acid sequence is given, written in the 5' → 3' direction. There are three types of oligonucleotides: PCR primer, fluorogenic probe used in the 5' nuclease assay, and complement used to hybridize to the corresponding probe. For the probes, the underlined base indicates a position where LAN with TAMRA attached was substituted for a T. (p) The presence of a 3' phosphate on each probe.

# Research

A1-2 RAQGCCCTCCCCATGCCATCCTGCGTp  
 A1-7 RATGCCCQCCCCATGCCATCCTGCGTp  
 A1-14 RATGCCCTCCCCCAQGCCATCCTGCGTp  
 A1-19 RATGCCCTCCCCATGCCAQCCTGCGTp  
 A1-22 RATGCCCTCCCCATGCCATCCQCGTp  
 A1-26 RATGCCCTCCCCATGCCATCCTGCGQp

Probe	518 nm		582 nm		RQ <sup>-</sup>	RQ <sup>+</sup>	ΔRQ
	no temp.	+ temp.	no temp.	+ temp.			
A1-2	25.5 ± 2.1	32.7 ± 1.9	38.2 ± 3.0	38.2 ± 2.0	0.67 ± 0.01	0.86 ± 0.06	0.19 ± 0.06
A1-7	53.5 ± 4.3	395.1 ± 21.4	108.5 ± 6.3	110.3 ± 5.3	0.49 ± 0.03	3.58 ± 0.17	3.09 ± 0.18
A1-14	127.0 ± 4.9	403.5 ± 19.1	109.7 ± 5.3	93.1 ± 6.3	1.16 ± 0.02	4.34 ± 0.15	3.18 ± 0.15
A1-19	187.5 ± 17.9	422.7 ± 7.7	70.3 ± 7.4	73.0 ± 2.8	2.67 ± 0.05	5.80 ± 0.15	3.13 ± 0.16
A1-22	224.6 ± 9.4	482.2 ± 43.6	100.0 ± 4.0	96.2 ± 9.6	2.25 ± 0.03	5.02 ± 0.11	2.77 ± 0.12
A1-26	160.2 ± 8.9	454.1 ± 18.4	93.1 ± 5.4	90.7 ± 3.2	1.72 ± 0.02	5.01 ± 0.08	3.29 ± 0.08

**FIGURE 2** Results of 5' nuclease assay comparing β-actin probes with TAMRA at different nucleotide positions. As described in Materials and Methods, PCR amplifications containing the indicated probes were performed, and the fluorescence emission was measured at 518 and 582 nm. Reported values are the average ± 1 S.D. for six reactions run without added template (no temp.) and six reactions run with template (+ temp.). The RQ ratio was calculated for each individual reaction and averaged to give the reported RQ<sup>-</sup> and RQ<sup>+</sup> values.

divided by the emission intensity of the quencher to give an RQ ratio for each reaction tube. This normalizes for well-to-well variations in probe concentration and fluorescence measurement. Finally, ΔRQ is calculated by subtracting the RQ value of the no-template control (RQ<sup>-</sup>) from the RQ value for the complete reaction including template (RQ<sup>+</sup>).

## RESULTS

A series of probes with increasing distances between the fluorescein reporter and rhodamine quencher were tested to investigate the minimum and maximum spacing that would give an acceptable performance in the 5' nuclease PCR assay. These probes hybridize to a target

sequence in the human β-actin gene. Figure 2 shows the results of an experiment in which these probes were included in PCR that amplified a segment of the β-actin gene containing the target sequence. Performance in the 5' nuclease PCR assay is monitored by the magnitude of ΔRQ, which is a measure of the increase in reporter fluorescence caused by PCR amplification of the probe target. Probe A1-2 has a ΔRQ value that is close to zero, indicating that the probe was not cleaved appreciably during the amplification reaction. This suggests that with the quencher dye on the second nucleotide from the 5' end, there is insufficient room for *Taq* polymerase to cleave efficiently between the reporter and quencher. The other five probes exhibited comparable ΔRQ values that are

clearly different from zero. Thus, all five probes are being cleaved during PCR amplification resulting in a similar increase in reporter fluorescence. It should be noted that complete digestion of a probe produces a much larger increase in reporter fluorescence than that observed in Figure 2 (data not shown). Thus, even in reactions where amplification occurs, the majority of probe molecules remain uncleaved. It is mainly for this reason that the fluorescence intensity of the quencher dye TAMRA changes little with amplification of the target. This is what allows us to use the 582-nm fluorescence reading as a normalization factor.

The magnitude of RQ<sup>-</sup> depends mainly on the quenching efficiency inherent in the specific structure of the probe and the purity of the oligonucleotide. Thus, the larger RQ<sup>-</sup> values indicate that probes A1-14, A1-19, A1-22, and A1-26 probably have reduced quenching as compared with A1-7. Still, the degree of quenching is sufficient to detect a highly significant increase in reporter fluorescence when each of these probes is cleaved during PCR.

To further investigate the ability of TAMRA on the 3' end to quench 6-FAM on the 5' end, three additional pairs of probes were tested in the 5' nuclease PCR assay. For each pair, one probe has TAMRA attached to an internal nucleotide and the other has TAMRA attached to the 3' end nucleotide. The results are shown in Table 2. For all three sets, the probe with the 3' quencher exhibits a ΔRQ value that is considerably higher than for the probe with the internal quencher. The RQ<sup>-</sup> values suggest that differences in quenching are not as great as those observed with some of the A1 probes. These results demonstrate that a quencher dye on the 3' end of an oligonucleotide can quench efficiently the

**TABLE 2** Results of 5' Nuclease Assay Comparing Probes with TAMRA Attached to an Internal or 3'-terminal Nucleotide

Probe	518 nm		582 nm		RQ <sup>-</sup>	RQ <sup>+</sup>	ΔRQ
	no temp.	+ temp.	no temp.	+ temp.			
A3-6	54.6 ± 3.2	84.8 ± 3.7	116.2 ± 6.4	115.6 ± 2.5	0.47 ± 0.02	0.73 ± 0.03	0.26 ± 0.04
A3-24	72.1 ± 2.9	236.5 ± 11.1	84.2 ± 4.0	90.2 ± 3.8	0.86 ± 0.02	2.62 ± 0.05	1.76 ± 0.05
P2-7	82.8 ± 4.4	384.0 ± 34.1	105.1 ± 6.4	120.4 ± 10.2	0.79 ± 0.02	3.19 ± 0.16	2.40 ± 0.16
P2-27	113.4 ± 6.6	555.4 ± 14.1	140.7 ± 8.5	118.7 ± 4.8	0.81 ± 0.01	4.68 ± 0.10	3.88 ± 0.10
P5-10	77.5 ± 6.5	244.4 ± 15.9	86.7 ± 4.3	95.8 ± 6.7	0.89 ± 0.05	2.55 ± 0.06	1.66 ± 0.08
P5-28	64.0 ± 5.2	333.6 ± 12.1	100.6 ± 6.1	94.7 ± 6.3	0.63 ± 0.02	3.53 ± 0.12	2.89 ± 0.13

Reactions containing the indicated probes and calculations were performed as described in Material and Methods and in the legend to Fig. 2.

## Research

fluorescence of a reporter dye on the 5' end. The degree of quenching is sufficient for this type of oligonucleotide to be used as a probe in the 5' nuclease PCR assay.

To test the hypothesis that quenching by a 3' TAMRA depends on the flexibility of the oligonucleotide, fluorescence was measured for probes in the single-stranded and double-stranded states. Table 3 reports the fluorescence observed at 518 and 582 nm. The relative degree of quenching is assessed by calculating the RQ ratio. For probes with TAMRA 6–10 nucleotides from the 5' end, there is little difference in the RQ values when comparing single-stranded with double-stranded oligonucleotides. The results for probes with TAMRA at the 3' end are much different. For these probes, hybridization to a complementary strand causes a dramatic increase in RQ. We propose that this loss of quenching is caused by the rigid structure of double-stranded DNA, which prevents the 5' and 3' ends from being in proximity.

When TAMRA is placed toward the 3' end, there is a marked  $Mg^{2+}$  effect on quenching. Figure 3 shows a plot of observed RQ values for the A1 series of probes as a function of  $Mg^{2+}$  concentration. With TAMRA attached near the 5' end (probe A1-2 or A1-7), the RQ value at 0 mM  $Mg^{2+}$  is only slightly higher than RQ at 10 mM  $Mg^{2+}$ . For probes A1-19, A1-22, and A1-26, the RQ values at 0 mM  $Mg^{2+}$  are very high, indicating a much

reduced quenching efficiency. For each of these probes, there is a marked decrease in RQ at 1 mM  $Mg^{2+}$  followed by a gradual decline as the  $Mg^{2+}$  concentration increases to 10 mM. Probe A1-14 shows an intermediate RQ value at 0 mM  $Mg^{2+}$  with a gradual decline at higher  $Mg^{2+}$  concentrations. In a low-salt environment with no  $Mg^{2+}$  present, a single-stranded oligonucleotide would be expected to adopt an extended conformation because of electrostatic repulsion. The binding of  $Mg^{2+}$  ions acts to shield the negative charge of the phosphate backbone so that the oligonucleotide can adopt conformations where the 3' end is close to the 5' end. Therefore, the observed  $Mg^{2+}$  effects support the notion that quenching of a 5' reporter dye by TAMRA at or near the 3' end depends on the flexibility of the oligonucleotide.

### DISCUSSION

The striking finding of this study is that it seems the rhodamine dye TAMRA, placed at any position in an oligonucleotide, can quench the fluorescent emission of a fluorescein (6-FAM) placed at the 5' end. This implies that a single-stranded, double-labeled oligonucleotide must be able to adopt conformations where the TAMRA is close to the 5' end. It should be noted that the decay of 6-FAM in the excited state requires a certain amount of time. Therefore, what

matters for quenching is not the average distance between 6-FAM and TAMRA but, rather, how close TAMRA can get to 6-FAM during the lifetime of the 6-FAM excited state. As long as the decay time of the excited state is relatively long compared with the molecular motions of the oligonucleotide, quenching can occur. Thus, we propose that TAMRA at the 3' end, or any other position, can quench 6-FAM at the 5' end because TAMRA is in proximity to 6-FAM often enough to be able to accept energy transfer from an excited 6-FAM.

Details of the fluorescence measurements remain puzzling. For example, Table 3 shows that hybridization of probes A1-26, A3-24, and P5-28 to their complementary strands not only causes a large increase in 6-FAM fluorescence at 518 nm but also causes a modest increase in TAMRA fluorescence at 582 nm. If TAMRA is being excited by energy transfer from quenched 6-FAM, then loss of quenching attributable to hybridization should cause a decrease in the fluorescence emission of TAMRA. The fact that the fluorescence emission of TAMRA increases indicates that the situation is more complex. For example, we have anecdotal evidence that the bases of the oligonucleotide, especially G, quench the fluorescence of both 6-FAM and TAMRA to some degree. When double-stranded, base-pairing may reduce the ability of the bases to quench. The primary factor causing the quenching of 6-FAM in an intact probe is the TAMRA dye. Evidence for the importance of TAMRA is that 6-FAM fluorescence remains relatively unchanged when probes labeled only with 6-FAM are used in the 5' nuclease PCR assay (data not shown). Secondary effectors of fluorescence, both before and after cleavage of the probe, need to be explored further.

Regardless of the physical mechanism, the relative independence of position and quenching greatly simplifies the design of probes for the 5' nuclease PCR assay. There are three main factors that determine the performance of a double-labeled fluorescent probe in the 5' nuclease PCR assay. The first factor is the degree of quenching observed in the intact probe. This is characterized by the value of  $RQ^{-}$ , which is the ratio of reporter to quencher fluorescent emissions for a no template control PCR. Influences on the value of  $RQ^{-}$  include the particular reporter and quencher

**TABLE 3** Comparison of Fluorescence Emissions of Single-stranded and Double-stranded Fluorogenic Probes

Probe	518 nm		582 nm		RQ	
	ss	ds	ss	ds	ss	ds
A1-7	27.75	68.53	61.08	138.18	0.45	0.50
A1-26	43.31	509.38	53.50	93.86	0.81	5.43
A3-6	16.75	62.88	39.33	165.57	0.43	0.38
A3-24	30.05	578.64	67.72	140.25	0.45	3.21
P2-7	35.02	70.13	54.63	121.09	0.64	0.58
P2-27	39.89	320.47	65.10	61.13	0.61	5.25
P5-10	27.34	144.85	61.95	165.54	0.44	0.87
P5-28	33.65	462.29	72.39	104.61	0.46	4.43

(ss) Single-stranded. The fluorescence emissions at 518 or 582 nm for solutions containing a final concentration of 50 nM indicated probe, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, and 10 mM  $MgCl_2$ . (ds) Double-stranded. The solutions contained, in addition, 100 nM A1C for probes A1-7 and A1-26, 100 nM A3C for probes A3-6 and A3-24, 100 nM P2C for probes P2-7 and P2-27, or 100 nM P5C for probes P5-10 and P5-28. Before the addition of  $MgCl_2$ , 120  $\mu$ l of each sample was heated at 95°C for 5 min. Following the addition of 80  $\mu$ l of 25 mM  $MgCl_2$ , each sample was allowed to cool to room temperature and the fluorescence emissions were measured. Reported values are the average of three determinations.

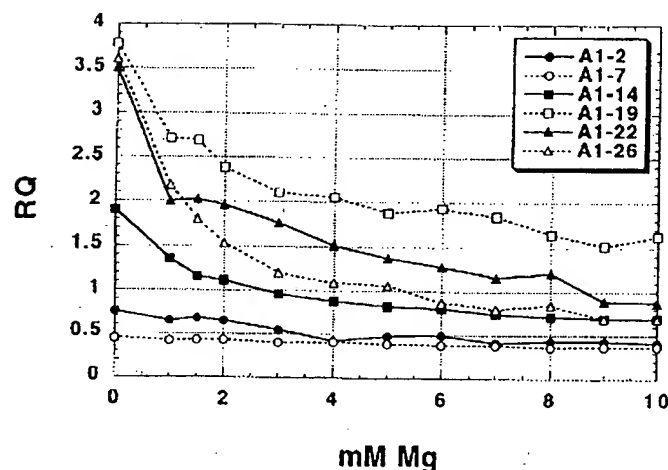


FIGURE 3 Effect of  $Mg^{2+}$  concentration on RQ ratio for the A1 series of probes. The fluorescence emission intensity at 518 and 582 nm was measured for solutions containing 50 nM probe, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, and varying amounts (0–10 mM) of  $MgCl_2$ . The calculated RQ ratios (518 nm intensity divided by 582 nm intensity) are plotted vs.  $MgCl_2$  concentration (mM Mg). The key (upper right) shows the probes examined.

dyes used, spacing between reporter and quencher dyes, nucleotide sequence context effects, presence of structure or other factors that reduce flexibility of the oligonucleotide, and purity of the probe. The second factor is the efficiency of hybridization, which depends on probe  $T_m$ , presence of secondary structure in probe or template, annealing temperature, and other reaction conditions. The third factor is the efficiency at which *Taq* DNA polymerase cleaves the bound probe between the reporter and quencher dyes. This cleavage is dependent on sequence complementarity between probe and template as shown by the observation that mismatches in the segment between reporter and quencher dyes drastically reduce the cleavage of probe.<sup>(1)</sup>

The rise in  $RQ^-$  values for the A1 series of probes seems to indicate that the degree of quenching is reduced somewhat as the quencher is placed toward the 3' end. The lowest apparent quenching is observed for probe A1-19 (see Fig. 3) rather than for the probe where the TAMRA is at the 3' end (A1-26). This is understandable, as the conformation of the 3' end position would be expected to be less restricted than the conformation of an internal position. In effect, a quencher at the 3' end is freer to adopt conformations close to the 5' reporter dye than is an internally placed quencher. For the other three sets of

probes, the interpretation of  $RQ^-$  values is less clear-cut. The A3 probes show the same trend as A1, with the 3' TAMRA probe having a larger  $RQ^-$  than the internal TAMRA probe. For the P2 pair, both probes have about the same  $RQ^-$  value. For the P5 probes, the  $RQ^-$  for the 3' probe is less than for the internally labeled probe. Another factor that may explain some of the observed variation is that purity affects the  $RQ^-$  value. Although all probes are HPLC purified, a small amount of contamination with unquenched reporter can have a large effect on  $RQ^-$ .

Although there may be a modest effect on degree of quenching, the position of the quencher apparently can have a large effect on the efficiency of probe cleavage. The most drastic effect is observed with probe A1-2, where placement of the TAMRA on the second nucleotide reduces the efficiency of cleavage to almost zero. For the A3, P2, and P5 probes,  $\Delta RQ$  is much greater for the 3' TAMRA probes as compared with the internal TAMRA probes. This is explained most easily by assuming that probes with TAMRA at the 3' end are more likely to be cleaved between reporter and quencher than are probes with TAMRA attached internally. For the A1 probes, the cleavage efficiency of probe A1-7 must already be quite high, as  $\Delta RQ$  does not increase when the quencher is placed closer to the 3' end. This illus-

trates the importance of being able to use probes with a quencher on the 3' end in the 5' nuclease PCR assay. In this assay, an increase in the intensity of reporter fluorescence is observed only when the probe is cleaved between the reporter and quencher dyes. By placing the reporter and quencher dyes on the opposite ends of an oligonucleotide probe, any cleavage that occurs will be detected. When the quencher is attached to an internal nucleotide, sometimes the probe works well (A1-7) and other times not so well (A3-6). The relatively poor performance of probe A3-6 presumably means the probe is being cleaved 3' to the quencher rather than between the reporter and quencher. Therefore, the best chance of having a probe that reliably detects accumulation of PCR product in the 5' nuclease PCR assay is to use a probe with the reporter and quencher dyes on opposite ends.

Placing the quencher dye on the 3' end may also provide a slight benefit in terms of hybridization efficiency. The presence of a quencher attached to an internal nucleotide might be expected to disrupt base-pairing and reduce the  $T_m$  of a probe. In fact, a 2°C–3°C reduction in  $T_m$  has been observed for two probes with internally attached TAMRAs.<sup>(9)</sup> This disruptive effect would be minimized by placing the quencher at the 3' end. Thus, probes with 3' quenchers might exhibit slightly higher hybridization efficiencies than probes with internal quenchers.

The combination of increased cleavage and hybridization efficiencies means that probes with 3' quenchers probably will be more tolerant of mismatches between probe and target as compared with internally labeled probes. This tolerance of mismatches can be advantageous, as when trying to use a single probe to detect PCR-amplified products from samples of different species. Also, it means that cleavage of probe during PCR is less sensitive to alterations in annealing temperature or other reaction conditions. The one application where tolerance of mismatches may be a disadvantage is for allelic discrimination. Lee et al.<sup>(11)</sup> demonstrated that allele-specific probes were cleaved between reporter and quencher only when hybridized to a perfectly complementary target. This allowed them to distinguish the normal human cystic fibrosis allele from the  $\Delta F508$  mutant. Their probes had TAMRA attached to the seventh nucleotide from

## Research

the 5' end and were designed so that any mismatches were between the reporter and quencher. Increasing the distance between reporter and quencher would lessen the disruptive effect of mismatches and allow cleavage of the probe on the incorrect target. Thus, probes with a quencher attached to an internal nucleotide may still be useful for allelic discrimination.

In this study loss of quenching upon hybridization was used to show that quenching by a 3' TAMRA is dependent on the flexibility of a single-stranded oligonucleotide. The increase in reporter fluorescence intensity, though, could also be used to determine whether hybridization has occurred or not. Thus, oligonucleotides with reporter and quencher dyes attached at opposite ends should also be useful as hybridization probes. The ability to detect hybridization in real time means that these probes could be used to measure hybridization kinetics. Also, this type of probe could be used to develop homogeneous hybridization assays for diagnostics or other applications. Bagwell et al.<sup>(10)</sup> describe just this type of homogeneous assay where hybridization of a probe causes an increase in fluorescence caused by a loss of quenching. However, they utilized a complex probe design that requires adding nucleotides to both ends of the probe sequence to form two imperfect hairpins. The results presented here demonstrate that the simple addition of a reporter dye to one end of an oligonucleotide and a quencher dye to the other end generates a fluorogenic probe that can detect hybridization or PCR amplification.

### ACKNOWLEDGMENTS

We acknowledge Lincoln McBride of Perkin-Elmer for his support and encouragement on this project and Mitch Winnik of the University of Toronto for helpful discussions on time-resolved fluorescence.

### REFERENCES

1. Lee, L.G., C.R. Connell, and W. Bloch. 1993. Allelic discrimination by nick-translation PCR with fluorogenic probes. *Nucleic Acids Res.* **21**: 3761-3766.
2. Holland, P.M., R.D. Abramson, R. Watson, and D.H. Gelfand. 1991. Detection of specific polymerase chain reaction prod-

uct by utilizing the 5' to 3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. Natl. Acad. Sci.* **88**: 7276-7280.

3. Lyamichev, V., M.A.D. Brow, and J.E. Dahlberg. 1993. Structure-specific endonucleolytic cleavage of nucleic acids by eubacterial DNA polymerases. *Science* **260**: 778-783.
4. Förster, V.Th. 1948. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann. Phys. (Leipzig)* **2**: 55-75.
5. Lakowicz, J.R. 1983. Energy transfer. In *Principles of fluorescent spectroscopy*, pp. 303-339. Plenum Press, New York, NY.
6. Stryer, L. and R.P. Haugland. 1967. Energy transfer: A spectroscopic ruler. *Proc. Natl. Acad. Sci.* **58**: 719-726.
7. Nakajima-Iijima, S., H. Hamada, P. Reddy, and T. Kakunaga. 1985. Molecular structure of the human cytoplasmic beta-actin gene: Inter-species homology of sequences in the introns. *Proc. Natl. Acad. Sci.* **82**: 6133-6137.
8. du Breuil, R.M., J.M. Patel, and B.V. Mendelow. 1993. Quantitation of  $\beta$ -actin-specific mRNA transcripts using xeno-competitive PCR. *PCR Methods Applic.* **3**: 57-59.
9. Livak, K.J. (unpubl.).
10. Bagwell, C.B., M.E. Munson, R.L. Christensen, and E.J. Lovett. 1994. A new homogeneous assay system for specific nucleic acid sequences: Poly-dA and poly-A detection. *Nucleic Acids Res.* **22**: 2424-2425.

Received December 20, 1994; accepted in revised form March 6, 1995.



THIS MATERIAL MAY BE PROTECTED  
BY COPYRIGHT LAW (17 U.S. CODE)

## GENOMIC METHODS

# Real Time Quantitative PCR

Christian A. Heid,<sup>1</sup> Junko Stevens,<sup>2</sup> Kenneth J. Livak,<sup>2</sup> and  
P. Mickey Williams<sup>1,3</sup>

<sup>1</sup>BioAnalytical Technology Department, Genentech, Inc., South San Francisco, California 94080;

<sup>2</sup>Applied BioSystems Division of Perkin Elmer Corp., Foster City, California 94404

We have developed a novel "real time" quantitative PCR method. The method measures PCR product accumulation through a dual-labeled fluorogenic probe (i.e., TaqMan Probe). This method provides very accurate and reproducible quantitation of gene copies. Unlike other quantitative PCR methods, real-time PCR does not require post-PCR sample handling, preventing potential PCR product carry-over contamination and resulting in much faster and higher throughput assays. The real-time PCR method has a very large dynamic range of starting target molecule determination (at least five orders of magnitude). Real-time quantitative PCR is extremely accurate and less-labor-intensive than current quantitative PCR methods.

Quantitative nucleic acid sequence analysis has had an important role in many fields of biological research. Measurement of gene expression (RNA) has been used extensively in monitoring biological responses to various stimuli (Tan et al. 1994; Huang et al. 1995a,b; Prud'homme et al. 1995). Quantitative gene analysis (DNA) has been used to determine the genomic quantity of a particular gene, as in the case of the human *HER2* gene, which is amplified in ~30% of breast tumors (Slamon et al. 1987). Gene and genome quantitation (DNA and RNA) also have been used for analysis of human immunodeficiency virus (HIV) burden demonstrating changes in the levels of virus throughout the different phases of the disease (Connor et al. 1993; Platak et al. 1993b; Furtado et al. 1995).

Many methods have been described for the quantitative analysis of nucleic acid sequences (both for RNA and DNA; Southern 1975; Sharp et al. 1980; Thomas 1980). Recently, PCR has proven to be a powerful tool for quantitative nucleic acid analysis. PCR and reverse transcriptase (RT)-PCR have permitted the analysis of minimal starting quantities of nucleic acid (as little as one cell equivalent). This has made possible many experiments that could not have been performed with traditional methods. Although PCR has provided a powerful tool, it is imperative

that it be used properly for quantitation (Ragymaekers 1995). Many early reports of quantitative PCR and RT-PCR described quantitation of the PCR product but did not measure the initial target sequence quantity. It is essential to design proper controls for the quantitation of the initial target sequences (Pere 1992; Clementi et al. 1993).

Researchers have developed several methods of quantitative PCR and RT-PCR. One approach measures PCR product quantity in the log phase of the reaction before the plateau (Kellogg et al. 1990; Pang et al. 1990). This method requires that each sample has equal input amounts of nucleic acid and that each sample under analysis amplifies with identical efficiency up to the point of quantitative analysis. A gene sequence (contained in all samples at relatively constant quantities, such as  $\beta$ -actin) can be used for sample amplification efficiency normalization. Using conventional methods of PCR detection and quantitation (gel electrophoresis or plate capture hybridization), it is extremely laborious to assure that all samples are analyzed during the log phase of the reaction (for both the target gene and the normalization gene). Another method, quantitative competitive (QC)-PCR, has been developed and is used widely for PCR quantitation. QC-PCR relies on the inclusion of an internal control competitor in each reaction (Becker-Andre 1991; Platak et al. 1993a,b). The efficiency of each reaction is normalized to the internal competitor. A known amount of internal competitor can be

<sup>3</sup>Corresponding author.

## REAL TIME QUANTITATIVE PCR

added to each sample. To obtain relative quantitation, the unknown target PCR product is compared with the known competitor PCR product. Success of a quantitative competitive PCR assay relies on developing an internal control that amplifies with the same efficiency as the target molecule. The design of the competitor and the validation of amplification efficiencies require a dedicated effort. However, because QC-PCR does not require that PCR products be analyzed during the log phase of the amplification, it is the easier of the two methods to use.

Several detection systems are used for quantitative PCR and RT-PCR analysis: (1) agarose gels, (2) fluorescent labeling of PCR products and detection with laser-induced fluorescence using capillary electrophoresis (Fusco et al. 1995; Williams et al. 1996) or acrylamide gels, and (3) plate capture and sandwich probe hybridization (Mulder et al. 1994). Although these methods proved successful, each method requires post-PCR manipulations that add time to the analysis and may lead to laboratory contamination. The sample throughput of these methods is limited (with the exception of the plate capture approach), and, therefore, these methods are not well suited for uses demanding high sample throughput (i.e., screening of large numbers of biomolecules or analyzing samples for diagnostics or clinical trials).

Here we report the development of a novel assay for quantitative DNA analysis. The assay is based on the use of the 5' nuclease assay first described by Holland et al. (1991). The method uses the 5' nuclease activity of *Taq* polymerase to cleave a nonextendible hybridization probe during the extension phase of PCR. The approach uses dual-labeled fluorogenic hybridization probes (Lee et al. 1993; Bassler et al. 1995; Ljvak et al. 1995a,b). One fluorescent dye serves as a reporter (FAM (i.e., 6-carboxyfluorescein)) and its emission spectra is quenched by the second fluorescent dye, TAMRA (i.e., 6-carboxy-tetramethylrhodamine). The nuclease degradation of the hybridization probe releases the quenching of the FAM fluorescent emission, resulting in an increase in peak fluorescent emission at 518 nm. The use of a sequence detector (ABI Prism) allows measurement of fluorescent spectra of all 96 wells of the thermal cycler continuously during the PCR amplification. Therefore, the reactions are monitored in real time. The output data is described and quantitative analysis of input target DNA sequences is discussed below.

## RESULTS

## PCR Product Detection in Real Time

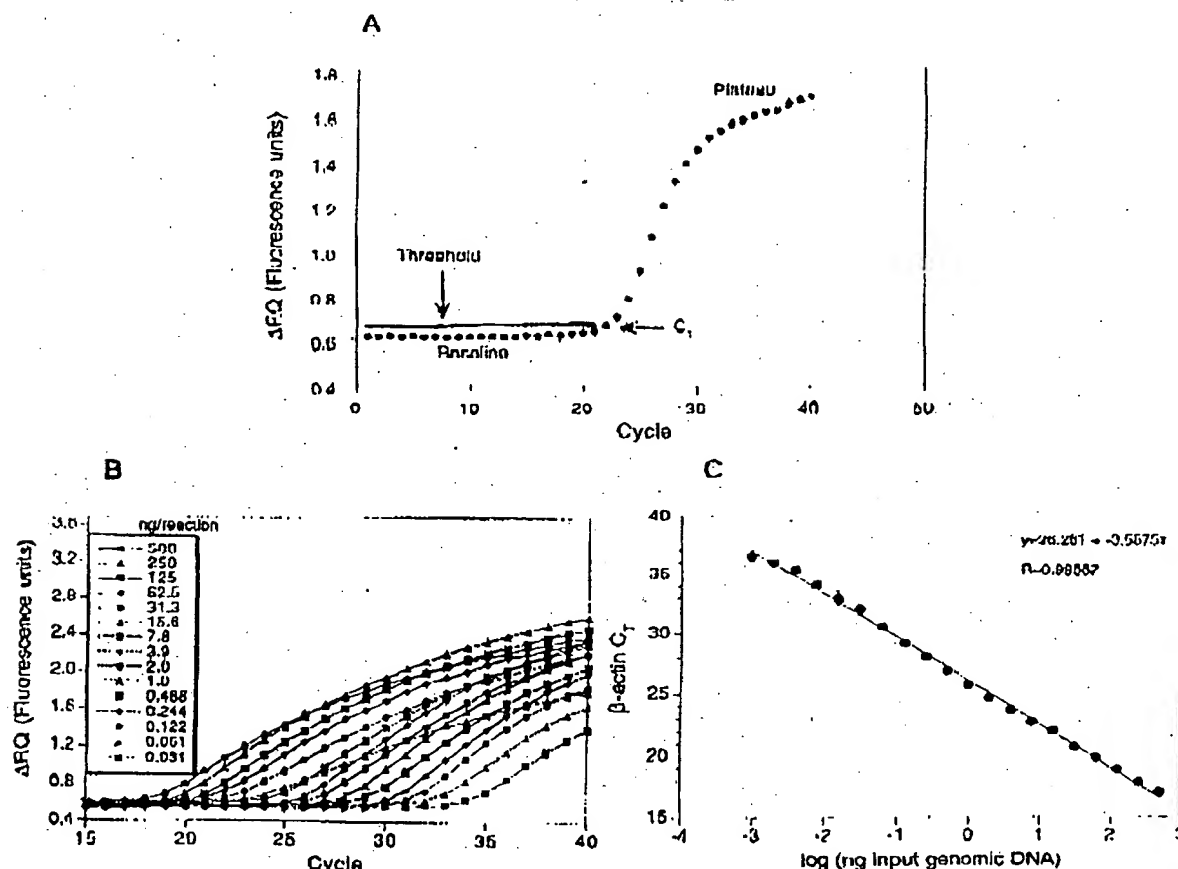
The goal was to develop a high-throughput, sensitive, and accurate gene quantitation assay for use in monitoring lipid mediated therapeutic gene delivery. A plasmid encoding human factor VIII gene sequence, pF8TM (see Methods), was used as a model therapeutic gene. The assay uses fluorescent Taqman methodology and an instrument capable of measuring fluorescence in real time (ABI Prism 7700 Sequence Detector). The Taqman reaction requires a hybridization probe labeled with two different fluorescent dyes. One dye is a reporter dye (FAM), the other is a quenching dye (TAMRA). When the probe is intact, fluorescent energy transfer occurs and the reporter dye fluorescent emission is absorbed by the quenching dye (TAMRA). During the extension phase of the PCR cycle, the fluorescent hybridization probe is cleaved by the 5'-3' nucleolytic activity of the DNA polymerase. On cleavage of the probe, the reporter dye emission is no longer transferred efficiently to the quenching dye, resulting in an increase of the reporter dye fluorescent emission spectra. PCR primers and probes were designed for the human factor VIII sequence and human  $\beta$ -actin gene (as described in Methods). Optimization reactions were performed to choose the appropriate probe and magnesium concentrations yielding the highest intensity of reporter fluorescent signal without sacrificing specificity. The instrument uses a charge-coupled device (i.e., CCD camera) for measuring the fluorescent emission spectra from 500 to 650 nm. Each PCR tube was monitored sequentially for 25 msec with continuous monitoring throughout the amplification. Each tube was re-examined every 8.5 sec. Computer software was designed to examine the fluorescent intensity of both the reporter dye (FAM) and the quenching dye (TAMRA). The fluorescent intensity of the quenching dye, TAMRA, changes very little over the course of the PCR amplification (data not shown). Therefore, the intensity of TAMRA dye emission serves as an internal standard with which to normalize the reporter dye (FAM) emission variations. The software calculates a value termed  $\Delta Rn$  (or  $\Delta R(2)$ ) using the following equation:  $\Delta Rn = (Rn^+) / (Rn^-)$ , where  $Rn^+$  = emission intensity of reporter/emission intensity of quencher at any given time in a reaction tube, and  $Rn^-$  = emission intensity of re-



## HUI ET AL.

porter/emission intensity of quencher measured prior to PCR amplification in that same reaction tube. For the purpose of quantitation, the last three data points ( $\Delta Rn$ s) collected during the extension step for each PCR cycle were analyzed. The nucleolytic degradation of the hybridization probe occurs during the extension phase of PCR, and, therefore, reporter fluorescent emission increases during this time. The three data points were averaged for each PCR cycle and the mean value for each was plotted in an "amplification plot" shown in Figure 1A. The  $\Delta Rn$  mean value is plotted on the y-axis, and time, represented by cycle number, is plotted on the x-axis. During the early cycles of the PCR amplification, the  $\Delta Rn$

value remains at base line. When sufficient hybridization probe has been cleaved by the *Taq* polymerase nuclease activity, the intensity of reporter fluorescent emission increases. Most PCR amplifications reach a plateau phase of reporter fluorescent emission if the reaction is carried out to high cycle numbers. The amplification plot is examined early in the reaction, at a point that represents the log phase of product accumulation. This is done by assigning an arbitrary threshold that is based on the variability of the base-line data. In Figure 1A, the threshold was set at 10 standard deviations above the mean of base line emission calculated from cycles 1 to 15. Once the threshold is chosen, the point at which



**Figure 1** PCR product detection in real time. (A) The Model 7700 software will construct amplification plots from the extension phase fluorescent emission data collected during the PCR amplification. The standard deviation is determined from the data points collected from the base line of the amplification plot.  $C_T$  values are calculated by determining the point at which the fluorescence exceeds a threshold limit (usually 10 times the standard deviation of the base line). (B) Overlay of amplification plots of serially (1:2) diluted human genomic DNA samples amplified with  $\beta$ -actin primers. (C) Input DNA concentration of the samples plotted versus  $C_T$ . All

## REAL TIME QUANTITATIVE PCR

the amplification plot crosses the threshold is defined as  $C_T$ .  $C_T$  is reported as the cycle number at this point. As will be demonstrated, the  $C_T$  value is predictive of the quantity of input target.

### $C_T$ Values Provide a Quantitative Measurement of Input Target Sequences

Figure 1B shows amplification plots of 15 different PCR amplifications overlaid. The amplifications were performed on a 1:2 serial dilution of human genomic DNA. The amplified target was human  $\beta$  actin. The amplification plots shift to the right (to higher threshold cycles) as the input target quantity is reduced. This is expected because reactions with fewer starting copies of the target molecule require greater amplification to degrade enough probe to attain the threshold fluorescence. An arbitrary threshold of 10 standard deviations above the base line was used to determine the  $C_T$  values. Figure 1C represents the  $C_T$  values plotted versus the sample dilution value. Each dilution was amplified in triplicate PCR amplifications and plotted as mean values with error bars representing one standard deviation. The  $C_T$  values decrease linearly with increasing target quantity. Thus,  $C_T$  values can be used as a quantitative measurement of the input target number. It should be noted that the amplification plot for the 15.6-ng sample shown in Figure 1B does not reflect the same fluorescent rate of increase exhibited by most of the other samples. The 15.6-ng sample also achieves endpoint plateau at a lower fluorescent value than would be expected based on the input DNA. This phenomenon has been observed occasionally with other samples (data not shown) and may be attributable to late cycle inhibition; this hypothesis is still under investigation. It is important to note that the flattened slope and early plateau do not impact significantly the calculated  $C_T$  value as demonstrated by the fit on the line shown in Figure 1C. All triplicate amplifications resulted in very similar  $C_T$  values—the standard deviation did not exceed 0.5 for any dilution. This experiment contains a >100,000-fold range of input target molecules. Using  $C_T$  values for quantitation permits a much larger assay range than directly using total fluorescent emission intensity for quantitation. The linear range of fluorescent intensity measurement of the ABI Prism 7700 Se-

ments over a very large range of relative starting target quantities.

### Sample Preparation Validation

Several parameters influence the efficiency of PCR amplification: magnesium and salt concentrations, reaction conditions (i.e., time and temperature), PCR target size and composition, primer sequences, and sample purity. All of the above factors are common to a single PCR assay, except sample to sample purity. In an effort to validate the method of sample preparation for the factor VIII assay, PCR amplification reproducibility and efficiency of 10 replicate sample preparations were examined. After genomic DNA was prepared from the 10 replicate samples, the DNA was quantitated by ultraviolet spectroscopy. Amplifications were performed analyzing  $\beta$ -actin gene content in 100 and 25 ng of total genomic DNA. Each PCR amplification was performed in triplicate. Comparison of  $C_T$  values for each triplicate sample show minimal variation based on standard deviation and coefficient of variance (Table 1). Therefore, each of the triplicate PCR amplifications was highly reproducible, demonstrating that real time PCR using this instrumentation introduces minimal variation into the quantitative PCR analysis. Comparison of the mean  $C_T$  values of the 10 replicate sample preparations also showed minimal variability, indicating that each sample preparation yielded similar results for  $\beta$ -actin gene quantity. The highest  $C_T$  difference between any of the samples was 0.85 and 0.71 for the 100 and 25 ng samples, respectively. Additionally, the amplification of each sample exhibited an equivalent rate of fluorescent emission intensity change per amount of DNA target analyzed as indicated by similar slopes derived from the sample dilutions (Fig. 2). Any sample containing an excess of a PCR inhibitor would exhibit a greater measured  $\beta$ -actin  $C_T$  value for a given quantity of DNA. In addition, the inhibitor would be diluted along with the sample in the dilution analysis (Fig. 2), altering the expected  $C_T$  value change. Each sample amplification yielded a similar result in the analysis, demonstrating that this method of sample preparation is highly reproducible with regard to sample purity.

### Quantitative Analysis of a Plasmid After

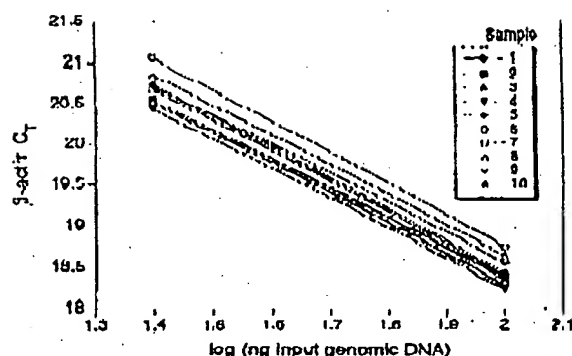
## III) FINAL

Table 1. Reproducibility of Sample Preparation Method

Sample no.	100 ng				25 ng			
	C <sub>T</sub>	mean	standard deviation	CV	C <sub>T</sub>	mean	standard deviation	CV
1	18.24	18.27	0.06	0.32	20.48	20.51	0.03	0.17
	18.23				20.55			
	18.33				20.5			
2	18.33	18.37	0.06	0.32	20.61	20.54	0.11	0.51
	18.35				20.59			
	18.44				20.41			
3	18.3	18.34	0.07	0.36	20.54	20.54	0.06	0.28
	18.3				20.6			
	18.42				20.49			
4	18.15	18.23	0.08	0.46	20.48	20.43	0.05	0.26
	18.23				20.44			
	18.32				20.38			
5	18.4	18.42	0.04	0.23	20.68	20.73	0.13	0.61
	18.38				20.87			
	18.46				20.63			
6	18.54	18.71	0.21	1.26	21.09	21.06	0.03	0.15
	18.67				21.04			
	19				21.01			
7	18.28	18.39	0.12	0.66	20.67	20.68	0.04	0.2
	18.36				20.73			
	18.52				20.65			
8	18.45	18.63	0.16	0.83	20.98	20.86	0.12	0.57
	18.7				20.84			
	18.73				20.75			
9	18.18	18.29	0.1	0.55	20.46	20.51	0.07	0.32
	18.34				20.54			
	18.26				20.48			
10	18.42	18.55	0.12	0.65	20.79	20.73	0.1	0.16
	18.57				20.78			
	18.66				20.62			
Mean	(1 10)	18.42	0.17	0.90		20.66	0.19	0.94

for containing a partial cDNA for human factor VIII, pF8TM. A series of transfections was set up using a decreasing amount of the plasmid (40, 4, 0.5, and 0.1 µg). Twenty-four hours post-transfection, total DNA was purified from each flask of cells. β-Actin gene quantity was chosen as a value for normalization of genomic DNA concentration from each sample. In this experiment, β-actin gene content should remain constant relative to total genomic DNA. Figure 3 shows the result of the β-actin DNA measurement (100 ng total DNA determined by ultraviolet spectroscopy) of each sample. Each sample was analyzed in triplicate and the mean β-actin C<sub>T</sub> values of the triplicates were plotted (error bars represent one standard deviation). The highest difference

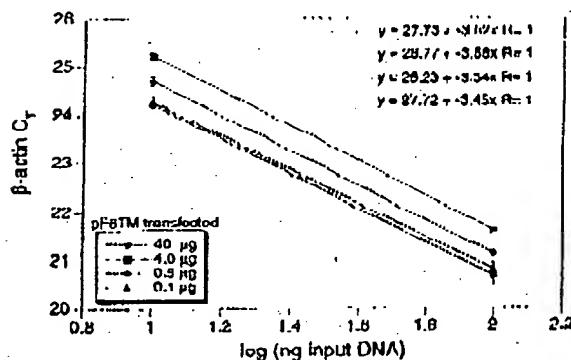
between any two sample means was 0.95 C<sub>T</sub>. Ten nanograms of total DNA of each sample were also examined for β-actin. The results again showed that very similar amounts of genomic DNA were present; the maximum mean β-actin C<sub>T</sub> value difference was 1.0. As Figure 3 shows, the rate of β-actin C<sub>T</sub> change between the 100 and 10-ng samples was similar (slope values range between 3.56 and -3.45). This verifies again that the method of sample preparation yields samples of identical PCR integrity (i.e., no sample contained an excessive amount of a PCR inhibitor). However, these results indicate that each sample contained slight differences in the actual amount of genomic DNA analyzed. Determination of actual genomic DNA concentration was accomplished



**Figure 2** Sample preparation purity. The replicate samples shown in Table 1 were also amplified in triplicate using 25 ng of each DNA sample. The figure shows the input DNA concentration (100 and 25 ng) vs.  $C_T$ . In the figure, the 100 and 25 ng points for each sample are connected by a line.

by plotting the mean  $\beta$ -actin  $C_T$  value obtained for each 100-ng sample on a  $\beta$ -actin standard curve (shown in Fig. 4C). The actual genomic DNA concentration of each sample,  $a$ , was obtained by extrapolation to the x-axis.

Figure 4A shows the measured (i.e., non-normalized) quantities of factor VIII plasmid DNA (pF8TM) from each of the four transient cell transfections. Each reaction contained 100 ng of total sample DNA (as determined by UV spectroscopy). Each sample was analyzed in triplicate



**Figure 3** Analysis of transfected cell DNA quantity and purity. The DNA preparations of the four 293 cell transfections (40, 4, 0.5, and 0.1  $\mu$ g of pF8TM) were analyzed for the  $\beta$ -actin gene. 100 and 10 ng (determined by ultraviolet spectroscopy) of each sample were amplified in triplicate. For each amount of pF8TM that was transfected, the  $\beta$ -actin  $C_T$  values are plotted versus the total input DNA concentration.

## REAL TIME QUANTITATIVE PCR

PCR amplifications. As shown, pF8TM purified from the 293 cells decreases (mean  $C_T$  values increase) with decreasing amounts of plasmid transfected. The mean  $C_T$  values obtained for pF8TM in Figure 4A were plotted on a standard curve comprised of serially diluted pF8TM, shown in Figure 4B. The quantity of pF8TM,  $b$ , found in each of the four transfections was determined by extrapolation to the x-axis of the standard curve in Figure 4B. These uncorrected values,  $b$ , for pF8TM were normalized to determine the actual amount of pF8TM found per 100 ng of genomic DNA by using the equation:

$$\frac{b \times 100 \text{ ng}}{a} = \text{actual pF8TM copies per 100 ng of genomic DNA}$$

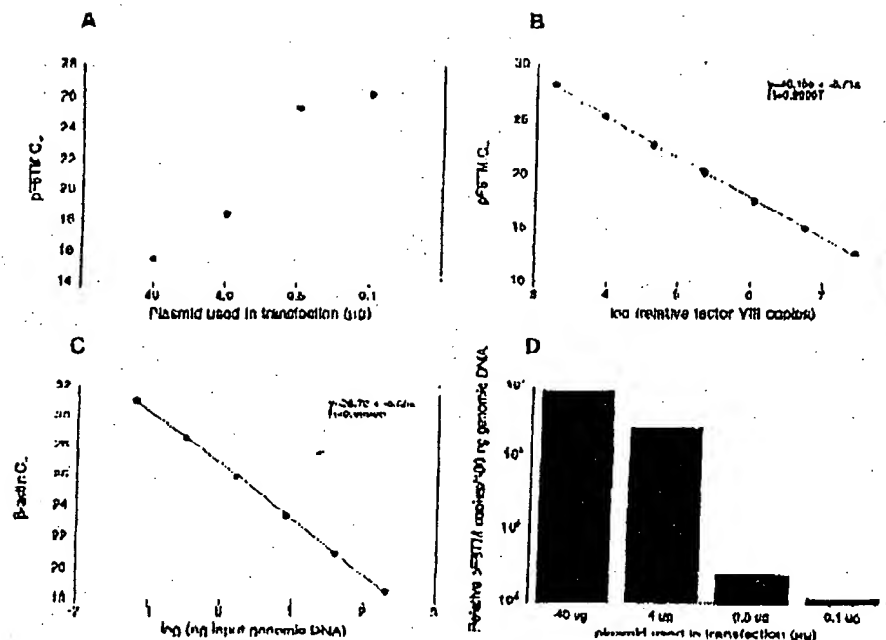
where  $a$  = actual genomic DNA in a sample and  $b$  = pF8TM copies from the standard curve. The normalized quantity of pF8TM per 100 ng of genomic DNA for each of the four transfections is shown in Figure 4D. These results show that the quantity of factor VIII plasmid associated with the 293 cells, 24 hr after transfection, decreases with decreasing plasmid concentration used in the transfection. The quantity of pF8TM associated with 293 cells, after transfection with 40  $\mu$ g of plasmid, was 35 pg per 100 ng genomic DNA. This results in ~520 plasmid copies per cell.

## DISCUSSION

We have described a new method for quantitating gene copy numbers using real-time analysis of PCR amplifications. Real-time PCR is compatible with either of the two PCR (RT-PCR) approaches: (1) quantitative competitive where an internal competitor for each target sequence is used for normalization (data not shown) or (2) quantitative comparative PCR using a normalization gene contained within the sample (i.e.,  $\beta$ -actin) or a "housekeeping" gene for RT-PCR. If equal amounts of nucleic acid are analyzed for each sample and if the amplification efficiency before quantitative analysis is identical for each sample, the internal control (normalization gene or competitor) should give equal signals for all samples.

The real-time PCR method offers several advantages over the other two methods currently employed (see the Introduction). First, the real-time PCR method is performed in a closed-tube system and requires no post-PCR manipulation

REID ET AL.



**Figure 4.** Quantitative analysis of pF8TM in transfected cells. (A) Amount of plasmid DNA used for the transfection plotted against the mean C<sub>T</sub> value determined for pF8TM remaining 24 hr after transfection. (B,C) Standard curves of pF8TM and β-actin, respectively. pF8TM DNA (B) and genomic DNA (C) were diluted serially 1:5 before amplification with the appropriate primers. The β-actin standard curve was used to normalize the results of A to 100 ng of genomic DNA. (D) The amount of pF8TM present per 100 ng of genomic DNA.

of sample. Therefore, the potential for PCR contamination in the laboratory is reduced because amplified products can be analyzed and disposed of without opening the reaction tubes. Second, this method supports the use of a normalization gene (i.e., β-actin) for quantitative PCR or house-keeping genes for quantitative RT-PCR controls. Analysis is performed in real time during the log phase of product accumulation. Analysis during log phase permits many different genes (over a wide input target range) to be analyzed simultaneously, without concern of reaching reaction plateau at different cycles. This will make multi-gene analysis assays much easier to develop, because individual internal competitors will not be needed for each gene under analysis. Third, sample throughput will increase dramatically with the new method because there is no post-PCR processing time. Additionally, working in a 96-well format is highly compatible with automation technology.

The real-time PCR method is highly reproducible. Replicate amplifications can be analyzed

for each sample minimizing potential error. The system allows for a very large assay dynamic range (approaching 1,000,000-fold starting target). Using a standard curve for the target of interest, relative copy number values can be determined for any unknown sample. Fluorescent threshold values, C<sub>T</sub>, correlate linearly with relative DNA copy numbers. Real time quantitative RT-PCR methodology (Gibson et al., this issue) has also been developed. Finally, real time quantitative PCR methodology can be used to develop high-throughput screening assays for a variety of applications [quantitative gene expression (RT-PCR), gene copy assays (Her2, HIV, etc.), genotyping (knockout mouse analysis), and immun-PCR].

Real-time PCR may also be performed using intercalating dyes (Higuchi et al. 1992) such as ethidium bromide. The fluorogenic probe method offers a major advantage over intercalating dyes—greater specificity (i.e., primer dimers and nonspecific PCR products are not detected).

## METHODS

### Generation of a Plasmid Containing a Partial cDNA for Human Factor VIII

Total RNA was harvested (RNAzol B from Tel Test, Inc., Friendswood, TX) from cells transfected with a factor VIII expression vector, pCIS2.8c251 (Eaton et al. 1986; Gorman et al. 1990). A factor VIII partial cDNA sequence was generated by RT-PCR (GeneAmp EZ RT/1h RNA PCR Kit (part N808-0179, PE Applied Biosystems, Foster City, CA)) using the PCR primers F8for and F8rev (primer sequences are shown below). The amplicon was reamplified using modified F8for and F8rev primers (appended with *hinf*II and *Hind*III restriction site sequences at the 5' end) and cloned into pGEM-3Z (Promega Corp., Madison, WI). The resulting clone, pF8TM, was used for transient transfection of 293 cells.

### Amplification of Target DNA and Detection of Amplicon Factor VIII Plasmid DNA

(pF8TM) was amplified with the primers F8for 5'-CCCCCTGCGCAAGAGCTGACCTTGTG-3' and F8rev 5'-AAACCTTCAAGCTGGATCGTACG-3'. The reaction produced a 422-bp PCR product. The forward primer was designed to recognize a unique sequence found in the 5' untranslated region of the parent pCIS2.8c251 plasmid and therefore does not recognize and amplify the human factor VIII gene. Primers were chosen with the assistance of the computer program Oligo 4.0 (National Biosciences, Inc., Plymouth, MN). The human  $\beta$ -actin gene was amplified with the primers  $\beta$ -actin forward primer 5'-TCACCCACACCTTCCCCATCTACGA-3' and  $\beta$ -actin reverse primer 5'-CAGCCGGAACCCGCTTCATTGCCAATGG-3'. The reaction produced a 295-bp PCR product.

Amplification reactions (50  $\mu$ l) contained a DNA sample, 10 $\times$  PCR Buffer II (5  $\mu$ l), 200  $\mu$ M dATP, dCTP, dGTP, and 400  $\mu$ M dUTP, 4 mM MgCl<sub>2</sub>, 1.25 Units AmpliTaq DNA polymerase, 0.5 unit AmpErase uracil N-glycosylase (UNG), 50 pmole of each factor VIII primer, and 15 pmole of each  $\beta$ -actin primer. The reactions also contained one of the following detection probes (100 nM each): F8probe 5'-(FAM)AGCTCTCTCCACCTGCTTCCTTCTGTGCTT(TAMRA)p-3' and  $\beta$ -actin probe 5'-(FAM)ATGCCCX(TAMRA)CCCCCATGCCATCp-3' where p indicates phosphorylation and X indicates a linker arm nucleotide. Reaction tubes were MicroAmp Optical Tubes (part number N801 0933, Perkin Elmer) that were frosted (at Perkin Elmer) to prevent light from reflecting. Tube caps were similar to MicroAmp Caps but specially designed to prevent light scattering. All of the PCR consumables were supplied by PE Applied Biosystems (Foster City, CA) except the factor VIII primers, which were synthesized at Genentech, Inc. (South San Francisco, CA). Probes were designed using the Oligo 4.0 software, following guidelines suggested in the Model 7700 Sequence Detector instrument manual. Briefly, probe T<sub>m</sub> should be at least 5°C higher than the annealing temperature used during thermal cycling; primers should not form stable duplexes with the probe.

The thermal cycling conditions included 2 min at 50°C and 10 min at 95°C. Thermal cycling proceeded with

## REAL TIME QUANTITATIVE PCR

reactions were performed in the Model 7700 Sequence Detector (PE Applied Biosystems), which contains a GeneAmp PCR System 9600. Reaction conditions were programmed on a Power Macintosh 7100 (Apple Computer, Santa Clara, CA) linked directly to the Model 7700 Sequence Detector. Analysis of data was also performed on the Macintosh computer. Collection and analysis software was developed at PE Applied Biosystems.

### Transfection of Cells with Factor VIII Construct

Four T175 flasks of 293 cells (ATCC CRL 1573), a human fetal kidney suspension cell line, were grown to 80% confluency and transfected pF8TM. Cells were grown in the following media: 50% HAM'S F12 without GHF, 50% low glucose Dulbecco's modified Eagle medium (DMEM) without glycine with sodium bicarbonate, 10% fetal bovine serum, 2 mM L-glutamine, and 1% penicillin-streptomycin. The media was changed 30 min before the transfection. pF8TM DNA amounts of 40, 4, 0.5, and 0.1  $\mu$ g were added to 1.5 ml of a solution containing 0.125 M CaCl<sub>2</sub> and 1 $\times$  HBEPIS. The four mixtures were left at room temperature for 10 min and then added dropwise to the cells. The flasks were incubated at 37°C and 5% CO<sub>2</sub> for 24 hr, washed with PBS, and resuspended in PBS. The resuspended cells were divided into aliquots and DNA was extracted immediately using the QIAamp Blood Kit (Qiagen, Chatsworth, CA). DNA was eluted into 200  $\mu$ l of 30 mM Tris-HCl at pH 8.0.

## ACKNOWLEDGMENTS

We thank Genentech's DNA Synthesis Group for primer synthesis and Genentech's Graphics Group for assistance with the figures.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

## REFERENCES

- Bassler, H.A., S.J. Flood, K.J. Livak, J. Marmaro, R. Kimm, and C.A. Ball. 1995. Use of a fluorogenic probe in a PCR-based assay for the detection of *Listeria monocytogenes*. *App. Environ. Microbiol.* 61: 3724-3720.
- Becker-Andre, M. 1991. Quantitative evaluation of mRNA levels. *Meth. Mol. Cell. Biol.* 2: 189-201.
- Clement, M., S. Menzo, P. Ingmirell, A. Manzi, A. Valenza, and P.E. Varaldo. 1993. Quantitative PCR and RT-PCR in virology. [Review]. *PCR Methods Applic.* 2: 191-196.
- Connor, R.I., H. Mohl, Y. Cao, and D.D. Ho. 1993. Increased viral burden and cytopathicity correlate temporally with CD4<sup>+</sup> T-lymphocyte decline and clinical progression in human immunodeficiency virus type 1-infected individuals. *J. Virol.* 67: 1772-1777.
- Eaton, D.L., W.J. Wood, D. Eaton, P.E. Hass, P.

## HFID L1 A1.

Venar, and C. Gorman. 1986. Construction and characterization of an active factor VIII variant lacking the central one third of the molecule. *Biochemistry* 25: 8343-8347.

Basco, M.J., C.P. Treanor, S. Spivack, H.L. Wigge, and L.S. Kaminsky. 1995. Quantitative RNA-polymerase chain reaction-DNA analysis by capillary electrophoresis and laser-induced fluorescence. *Anal. Biochem.* 224: 140-147.

Perre, J. 1992. Quantitative or semi-quantitative PCR: Reality versus myth. *PCR Methods Appl.* 2: 1-9.

Fortado, M.R., L.A. Kingsley, and S.M. Wollinsky. 1995. Changes in the viral mRNA expression pattern correlate with a rapid rate of CD4+ T-cell number decline in human immunodeficiency virus type 1-infected individuals. *J. Virol.* 69: 2092-2100.

Gibson, U.E.M., C.A. Heid, and P.M. Williams. 1996. A novel method for real time quantitative competitive RT-PCR. *Genome Res.* (this issue).

Gorman, C.M., D.R. Gies, and G. McCray. 1990. Transient production of proteins using an adenovirus transformed cell line. *DNA Prot. Engin. Tech.* 2: 3-10.

Higuchi, R., G. Dollinger, P.S. Walsh, and R. Griffith. 1992. Simultaneous amplification and detection of specific DNA sequences. *Biotechnology* 10: 413-417.

Holland, P.M., R.D. Abramson, R. Watson, and D.J. Gelfand. 1991. Detection of specific polymerase chain reaction product by utilizing the 5'-3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. Natl. Acad. Sci.* 88: 7276-7280.

Huang, S.K., H.Q. Xiao, T.J. Klein, G. Paciotti, H.G. Marsh, L.M. Lichtenstein, and M.C. Liu. 1995a. IL-13 expression at the sites of allergen challenge in patients with asthma. *J. Immun.* 155: 2688-2694.

Huang, S.K., M. Yi, E. Palmer, and D.G. Marsh. 1995b. A dominant T cell receptor beta-chain in response to a short ragweed allergen. *Am J. Immun.* 164: 6157-6162.

Kellogg, D.E., J.J. Shtinsky, and S. Kowk. 1990. Quantitation of HIV-1 proviral DNA relative to cellular DNA by the polymerase chain reaction. *Anal. Biochem.* 189: 202-208.

Lee, J.-G., C.R. Connell, and W. Bloch. 1993. Allelic discrimination by nick-translation PCR with fluorogenic probes. *Nucleic Acids Res.* 21: 3761-3766.

Livak, K.J., S.J. Flood, J. Marmaro, W. Giusti, and K. Deetz. 1995a. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods Appl.* 4: 357-362.

Livak, K.J., J. Marmaro, and J.A. Todd. 1995b. Towards

fully automated genome-wide polymorphism screening. [Letter] *Nature Genet.* 9: 341-342.

Mulder, J., N. McKinney, C. Christopherson, J. Shtinsky, L. Greenfield, and S. Kwok. 1994. Rapid and simple PCR assay for quantitation of human immunodeficiency virus type 1 RNA in plasma: Application to acute retroviral infection. *J. Clin. Microbiol.* 32: 292-300.

Pang, S., Y. Koyanagi, S. Miles, C. Wiley, H.V. Vinters, and L.S. Chen. 1990. High levels of unintegrated HIV-1 DNA in brain tissue of AIDS dementia patients. *Nature* 343: 85-89.

Platak, M.J., K.C. Luk, B. Williams, and J.D. Lifson. 1993a. Quantitative competitive polymerase chain reaction for accurate quantitation of HIV DNA and RNA species. *AltTechniques* 14: 70-81.

Platak, M.J., M.S. Saag, L.C. Yang, S.J. Clark, J.C. Kappes, K.C. Luk, B.H. Hann, G.M. Shaw, and J.D. Lifson. 1993b. High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR [see Comments]. *Science* 259: 1749-1754.

Prud'homme, G.J., D.H. Kono, and A.N. Theofilopoulos. 1995. Quantitative polymerase chain reaction analysis reveals marked overexpression of interleukin-1 beta, interleukin-1 and interferon-gamma mRNA in the lymph nodes of lupus-prone mice. *Mol. Immunol.* 32: 495-503.

Racymackers, L. 1995. A commentary on the practical applications of competitive PCR. *Genome Res.* 5: 91-94.

Sharp, P.A., A.J. Berk, and S.M. Berget. 1980. Transcription maps of adenovirus. *Methods Enzymol.* 65: 750-768.

Slamon, D.J., G.M. Clark, S.C. Wong, W.J. Levin, A. Ullrich, and W.J. McGuire. 1987. Human breast cancer: Correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* 235: 177-182.

Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98: 503-517.

Tan, X., X. Sun, C.F. Gonzalez, and W. Hsueh. 1994. pAI and TNF increase the precursor of Delta-kappa B p50 mRNA in mouse intestine: Quantitative analysis by competitive PCR. *Biochim. Biophys. Acta* 1215: 157-162.

Thomas, P.S. 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Natl. Acad. Sci.* 77: 5201-5205.

Williams, S., C. Schwer, A. Krishnamo, C. Held, B. Karger, and P.M. Williams. 1996. Quantitative competitive PCR: Analysis of amplified products of the HIV-1 gag gene by capillary electrophoresis with laser induced fluorescence detection. *Anal. Biochem.* (in press).

Received June 3, 1996; accepted in revised form July 29, 1996.

## WISP genes are members of the connective tissue growth factor family that are up-regulated in Wnt-1-transformed cells and aberrantly expressed in human colon tumors

DIANE PENNICA\*†, TODD A. SWANSON\*, JAMES W. WELSH\*, MARGARET A. ROY‡, DAVID A. LAWRENCE\*, JAMES LEE‡, JENNIFER BRUSH‡, LISA A. TANEYHILL§, BETHANNE DEUEL‡, MICHAEL LEW¶, COLIN WATANABE||, ROBERT L. COHEN\*, MONA F. MELHEM\*\*, GENE G. FINLEY\*\*, PHIL QUIRKE††, AUDREY D. GODDARD‡, KENNETH J. HILLAN¶, AUSTIN L. GURNEY‡, DAVID BOTSTEIN‡,††, AND ARNOLD J. LEVINE§

Departments of \*Molecular Oncology, ‡Molecular Biology, §Scientific Computing, and ¶Pathology, Genentech Inc., 1 DNA Way, South San Francisco, CA 94080; \*\*University of Pittsburgh School of Medicine, Veterans Administration Medical Center, Pittsburgh, PA 15240; ††University of Leeds, Leeds, LS29JT United Kingdom; ‡‡Department of Genetics, Stanford University, Palo Alto, CA 94305; and §Department of Molecular Biology, Princeton University, Princeton, NJ 08544

Contributed by David Botstein and Arnold J. Levine, October 21, 1998

**ABSTRACT** Wnt family members are critical to many developmental processes, and components of the Wnt signaling pathway have been linked to tumorigenesis in familial and sporadic colon carcinomas. Here we report the identification of two genes, *WISP-1* and *WISP-2*, that are up-regulated in the mouse mammary epithelial cell line C57MG transformed by Wnt-1, but not by Wnt-4. Together with a third related gene, *WISP-3*, these proteins define a subfamily of the connective tissue growth factor family. Two distinct systems demonstrated *WISP* induction to be associated with the expression of Wnt-1. These included (i) C57MG cells infected with a Wnt-1 retroviral vector or expressing Wnt-1 under the control of a tetracycline repressible promoter, and (ii) Wnt-1 transgenic mice. The *WISP-1* gene was localized to human chromosome 8q24.1-8q24.3. *WISP-1* genomic DNA was amplified in colon cancer cell lines and in human colon tumors and its RNA overexpressed (2- to >30-fold) in 84% of the tumors examined compared with patient-matched normal mucosa. *WISP-3* mapped to chromosome 6q22-6q23 and also was overexpressed (4- to >40-fold) in 63% of the colon tumors analyzed. In contrast, *WISP-2* mapped to human chromosome 20q12-20q13 and its DNA was amplified, but RNA expression was reduced (2- to >30-fold) in 79% of the tumors. These results suggest that the *WISP* genes may be downstream of Wnt-1 signaling and that aberrant levels of *WISP* expression in colon cancer may play a role in colon tumorigenesis.

Wnt-1 is a member of an expanding family of cysteine-rich, glycosylated signaling proteins that mediate diverse developmental processes such as the control of cell proliferation, adhesion, cell polarity, and the establishment of cell fates (1, 2). Wnt-1 originally was identified as an oncogene activated by the insertion of mouse mammary tumor virus in virus-induced mammary adenocarcinomas (3, 4). Although Wnt-1 is not expressed in the normal mammary gland, expression of Wnt-1 in transgenic mice causes mammary tumors (5).

In mammalian cells, Wnt family members initiate signaling by binding to the seven-transmembrane spanning Frizzled receptors and recruiting the cytoplasmic protein Dishevelled (Dsh) to the cell membrane (1, 2, 6). Dsh then inhibits the kinase activity of the normally constitutively active glycogen synthase kinase-3 $\beta$  (GSK-3 $\beta$ ) resulting in an increase in  $\beta$ -catenin levels. Stabilized  $\beta$ -catenin interacts with the transcription factor TCF/Lef1, forming a complex that appears in

the nucleus and binds TCF/Lef1 target DNA elements to activate transcription (7, 8). Other experiments suggest that the adenomatous polyposis coli (APC) tumor suppressor gene also plays an important role in Wnt signaling by regulating  $\beta$ -catenin levels (9). APC is phosphorylated by GSK-3 $\beta$ , binds to  $\beta$ -catenin, and facilitates its degradation. Mutations in either APC or  $\beta$ -catenin have been associated with colon carcinomas and melanomas, suggesting these mutations contribute to the development of these types of cancer, implicating the Wnt pathway in tumorigenesis (1).

Although much has been learned about the Wnt signaling pathway over the past several years, only a few of the transcriptionally activated downstream components activated by Wnt have been characterized. Those that have been described cannot account for all of the diverse functions attributed to Wnt signaling. Among the candidate Wnt target genes are those encoding the nodal-related 3 gene, *Xnr3*, a member of the transforming growth factor (TGF)- $\beta$  superfamily, and the homeobox genes, *engrailed*, *gooseoid*, *twin* (*Xtwn*), and *siamois* (2). A recent report also identifies *c-myc* as a target gene of the Wnt signaling pathway (10).

To identify additional downstream genes in the Wnt signaling pathway that are relevant to the transformed cell phenotype, we used a PCR-based cDNA subtraction strategy, suppression subtractive hybridization (SSH) (11), using RNA isolated from C57MG mouse mammary epithelial cells and C57MG cells stably transformed by a Wnt-1 retrovirus. Overexpression of Wnt-1 in this cell line is sufficient to induce a partially transformed phenotype, characterized by elongated and refractile cells that lose contact inhibition and form a multilayered array (12, 13). We reasoned that genes differentially expressed between these two cell lines might contribute to the transformed phenotype.

In this paper, we describe the cloning and characterization of two genes up-regulated in Wnt-1 transformed cells, *WISP-1* and *WISP-2*, and a third related gene, *WISP-3*. The *WISP* genes are members of the CCN family of growth factors, which includes connective tissue growth factor (CTGF), Cyr61, and *nov*, a family not previously linked to Wnt signaling.

### MATERIALS AND METHODS

**SSH.** SSH was performed by using the PCR-Select cDNA Subtraction Kit (CLONTECH). Tester double-stranded

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1998 by The National Academy of Sciences 0027-8424/98/9514717-6\$2.00/0 PNAS is available online at www.pnas.org.

Abbreviations: TGF, transforming growth factor; CTGF, connective tissue growth factor; SSH, suppression subtractive hybridization; WVC, von Willebrand factor type C module.

Data deposition: The sequences reported in this paper have been deposited in the Genbank database (accession nos. AF100777, AF100778, AF100779, AF100780, and AF100781).

†To whom reprint requests should be addressed. e-mail: diane@gene.com.



cDNA was synthesized from 2  $\mu$ g of poly(A)<sup>+</sup> RNA isolated from the C57MG/Wnt-1 cell line and driver cDNA from 2  $\mu$ g of poly(A)<sup>+</sup> RNA from the parent C57MG cells. The subtracted cDNA library was subcloned into a pGEM-T vector for further analysis.

**cDNA Library Screening.** Clones encoding full-length mouse *WISP-1* were isolated by screening a  $\lambda$ gt10 mouse embryo cDNA library (CLONTECH) with a 70-bp probe from the original partial clone 568 sequence corresponding to amino acids 128–169. Clones encoding full-length human *WISP-1* were isolated by screening  $\lambda$ gt10 lung and fetal kidney cDNA libraries with the same probe at low stringency. Clones encoding full-length mouse and human *WISP-2* were isolated by screening a C57MG/Wnt-1 or human fetal lung cDNA library with a probe corresponding to nucleotides 1463–1512. Full-length cDNAs encoding *WISP-3* were cloned from human bone marrow and fetal kidney libraries.

**Expression of Human *WISP* RNA.** PCR amplification of first-strand cDNA was performed with human Multiple Tissue cDNA panels (CLONTECH) and 300  $\mu$ M of each dNTP at 94°C for 1 sec, 62°C for 30 sec, 72°C for 1 min, for 22–32 cycles. *WISP* and glyceraldehyde-3-phosphate dehydrogenase primer sequences are available on request.

**In Situ Hybridization.** <sup>32</sup>P-labeled sense and antisense riboprobes were transcribed from an 897-bp PCR product corresponding to nucleotides 601–1440 of mouse *WISP-1* or a 294-bp PCR product corresponding to nucleotides 82–375 of mouse *WISP-2*. All tissues were processed as described (40).

**Radiation Hybrid Mapping.** Genomic DNA from each hybrid in the Stanford G3 and Genebridge4 Radiation Hybrid Panels (Research Genetics, Huntsville, AL) and human and hamster control DNAs were PCR-amplified, and the results were submitted to the Stanford or Massachusetts Institute of Technology web servers.

**Cell Lines, Tumors, and Mucosa Specimens.** Tissue specimens were obtained from the Department of Pathology (University of Pittsburgh) for patients undergoing colon resection and from the University of Leeds, United Kingdom. Genomic DNA was isolated (Qiagen) from the pooled blood of 10 normal human donors, surgical specimens, and the following ATCC human cell lines: SW480, COLO 320DM, HT-29, WiDr, and SW403 (colon adenocarcinomas), SW620 (lymph node metastasis, colon adenocarcinoma), HCT 116 (colon carcinoma), SK-CO-1 (colon adenocarcinoma, ascites), and HM7 (a variant of ATCC colon adenocarcinoma cell line LS 174T). DNA concentration was determined by using Hoechst dye 33258 intercalation fluorimetry. Total RNA was prepared by homogenization in 7 M GuSCN followed by centrifugation over CsCl cushions or prepared by using RNeasy.

**Gene Amplification and RNA Expression Analysis.** Relative gene amplification and RNA expression of *WISPs* and *c-myc* in the cell lines, colorectal tumors, and normal mucosa were determined by quantitative PCR. Gene-specific primers and fluorogenic probes (sequences available on request) were designed and used to amplify and quantitate the genes. The relative gene copy number was derived by using the formula  $2^{\Delta\Delta Ct}$  where  $\Delta Ct$  represents the difference in amplification cycles required to detect the *WISP* genes in peripheral blood lymphocyte DNA compared with colon tumor DNA or colon tumor RNA compared with normal mucosal RNA. The  $\Delta$ -method was used for calculation of the SE of the gene copy number or RNA expression level. The *WISP*-specific signal was normalized to that of the glyceraldehyde-3-phosphate dehydrogenase housekeeping gene. All TaqMan assay reagents were obtained from Perkin-Elmer Applied Biosystems.

## RESULTS

**Isolation of *WISP-1* and *WISP-2* by SSH.** To identify Wnt-1-inducible genes, we used the technique of SSH using the

mouse mammary epithelial cell line C57MG and C57MG cells that stably express Wnt-1 (11). Candidate differentially expressed cDNAs (1,384 total) were sequenced. Thirty-nine percent of the sequences matched known genes or homologues, 32% matched expressed sequence tags, and 29% had no match. To confirm that the transcript was differentially expressed, semiquantitative reverse transcription-PCR and Northern analysis were performed by using mRNA from the C57MG and C57MG/Wnt-1 cells.

Two of the cDNAs, *WISP-1* and *WISP-2*, were differentially expressed, being induced in the C57MG/Wnt-1 cell line, but not in the parent C57MG cells or C57MG cells overexpressing Wnt-4 (Fig. 1A and B). Wnt-4, unlike Wnt-1, does not induce the morphological transformation of C57MG cells and has no effect on  $\beta$ -catenin levels (13, 14). Expression of *WISP-1* was up-regulated approximately 3-fold in the C57MG/Wnt-1 cell line and *WISP-2* by approximately 5-fold by both Northern analysis and reverse transcription-PCR.

An independent, but similar, system was used to examine *WISP* expression after Wnt-1 induction. C57MG cells expressing the *Wnt-1* gene under the control of a tetracycline-repressible promoter produce low amounts of Wnt-1 in the repressed state but show a strong induction of *Wnt-1* mRNA and protein within 24 hr after tetracycline removal (8). The levels of Wnt-1 and *WISP* RNA isolated from these cells at various times after tetracycline removal were assessed by quantitative PCR. Strong induction of Wnt-1 mRNA was seen as early as 10 hr after tetracycline removal. Induction of *WISP* mRNA (2- to 6-fold) was seen at 48 and 72 hr (data not shown). These data support our previous observations that show that *WISP* induction is correlated with Wnt-1 expression. Because the induction is slow, occurring after approximately 48 hr, the induction of *WISPs* may be an indirect response to Wnt-1 signaling.

cDNA clones of human *WISP-1* were isolated and the sequence compared with mouse *WISP-1*. The cDNA sequences of mouse and human *WISP-1* were 1,766 and 2,830 bp in length, respectively, and encode proteins of 367 aa, with predicted relative molecular masses of  $\approx 40,000$  ( $M_r$ , 40 K). Both have hydrophobic N-terminal signal sequences, 38 conserved cysteine residues, and four potential N-linked glycosylation sites and are 84% identical (Fig. 2A).

Full-length cDNA clones of mouse and human *WISP-2* were 1,734 and 1,293 bp in length, respectively, and encode proteins of 251 and 250 aa, respectively, with predicted relative molecular masses of  $\approx 27,000$  ( $M_r$ , 27 K) (Fig. 2B). Mouse and human *WISP-2* are 73% identical. Human *WISP-2* has no potential N-linked glycosylation sites, and mouse *WISP-2* has one at

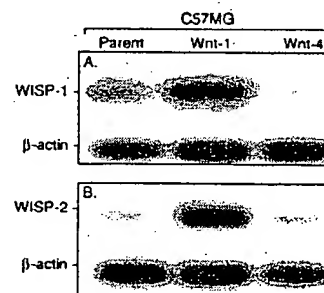


FIG. 1. *WISP-1* and *WISP-2* are induced by Wnt-1, but not Wnt-4, expression in C57MG cells. Northern analysis of *WISP-1* (A) and *WISP-2* (B) expression in C57MG, C57MG/Wnt-1, and C57MG/Wnt-4 cells. Poly(A)<sup>+</sup> RNA (2  $\mu$ g) was subjected to Northern blot analysis and hybridized with a 70-bp mouse *WISP-1*-specific probe (amino acids 278–300) or a 190-bp *WISP-2*-specific probe (nucleotides 1438–1627) in the 3' untranslated region. Blots were rehybridized with human  $\beta$ -actin probe.

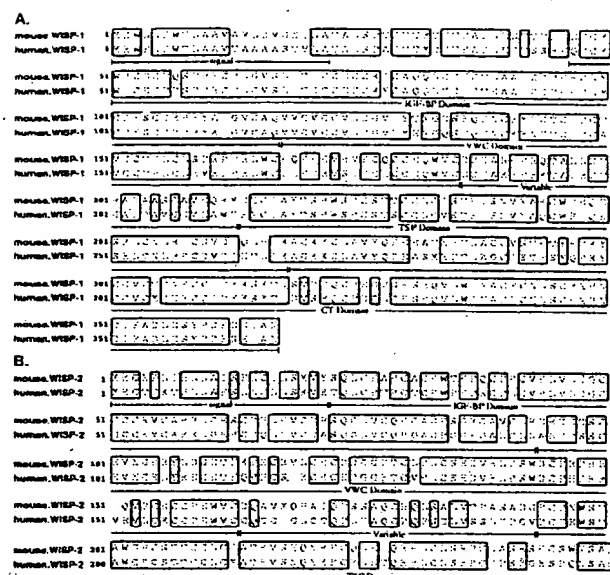


FIG. 2. Encoded amino acid sequence alignment of mouse and human *WISP-1* (A) and mouse and human *WISP-2* (B). The potential signal sequence, insulin-like growth factor-binding protein (IGF-BP), VWC, thrombospondin (TSP), and C-terminal (CT) domains are underlined.

position 197. *WISP-2* has 28 cysteine residues that are conserved among the 38 cysteines found in *WISP-1*.

**Identification of *WISP-3*.** To search for related proteins, we screened expressed sequence tag (EST) databases with the *WISP-1* protein sequence and identified several ESTs as potentially related sequences. We identified a homologous protein that we have called *WISP-3*. A full-length human *WISP-3* cDNA of 1,371 bp was isolated corresponding to those ESTs that encode a 354-aa protein with a predicted molecular mass of 39,293. *WISP-3* has two potential N-linked glycosylation sites and 36 cysteine residues. An alignment of the three human *WISP* proteins shows that *WISP-1* and *WISP-3* are the most similar (42% identity), whereas *WISP-2* has 37% identity with *WISP-1* and 32% identity with *WISP-3* (Fig. 3A).

***WISPs* Are Homologous to the CTGF Family of Proteins.** Human *WISP-1*, *WISP-2*, and *WISP-3* are novel sequences; however, mouse *WISP-1* is the same as the recently identified *Elm1* gene. *Elm1* is expressed in low, but not high, metastatic mouse melanoma cells, and suppresses the *in vivo* growth and metastatic potential of K-1735 mouse melanoma cells (15). Human and mouse *WISP-2* are homologous to the recently described rat gene, *rCop-1* (16). Significant homology (36–44%) was seen to the CCN family of growth factors. This family includes three members, CTGF, Cyr61, and the protooncogene *nov*. CTGF is a chemotactic and mitogenic factor for fibroblasts that is implicated in wound healing and fibrotic disorders and is induced by TGF- $\beta$  (17). Cyr61 is an extracellular matrix signaling molecule that promotes cell adhesion, proliferation, migration, angiogenesis, and tumor growth (18, 19). *nov* (nephroblastoma overexpressed) is an immediate early gene associated with quiescence and found altered in Wilms' tumors (20). The proteins of the CCN family share functional, but not sequence, similarity to Wnt-1. All are secreted, cysteine-rich heparin binding glycoproteins that associate with the cell surface and extracellular matrix.

*WISP* proteins exhibit the modular architecture of the CCN family, characterized by four conserved cysteine-rich domains (Fig. 3B) (21). The N-terminal domain, which includes the first 12 cysteine residues, contains a consensus sequence (GCGC-CXXC) conserved in most insulin-like growth factor (IGF)-

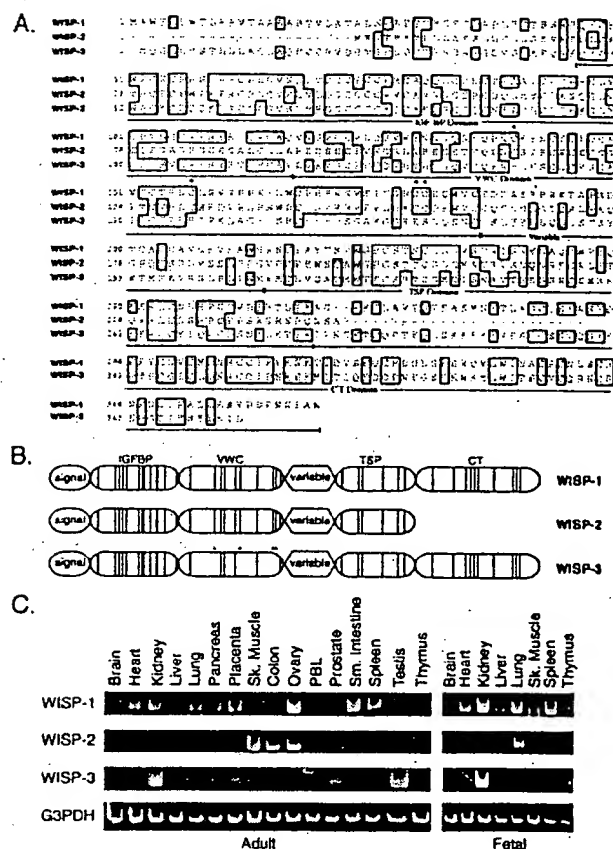


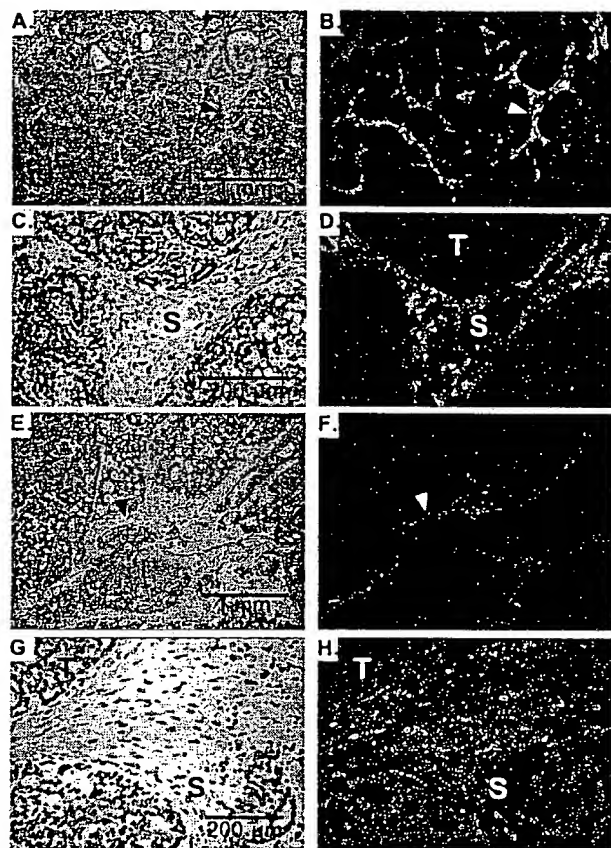
FIG. 3. (A) Encoded amino acid sequence alignment of human *WISPs*. The cysteine residues of *WISP-1* and *WISP-2* that are not present in *WISP-3* are indicated with a dot. (B) Schematic representation of the *WISP* proteins showing the domain structure and cysteine residues (vertical lines). The four cysteine residues in the VWC domain that are absent in *WISP-3* are indicated with a dot. (C) Expression of *WISP* mRNA in human tissues. PCR was performed on human multiple-tissue cDNA panels (CLONTECH) from the indicated adult and fetal tissues.

binding proteins (BP). This sequence is conserved in *WISP-2* and *WISP-3*, whereas *WISP-1* has a glutamine in the third position instead of a glycine. CTGF recently has been shown to specifically bind IGF (22) and a truncated *nov* protein lacking the IGF-BP domain is oncogenic (23). The von Willebrand factor type C module (VWC), also found in certain collagens and mucins, covers the next 10 cysteine residues, and is thought to participate in protein complex formation and oligomerization (24). The VWC domain of *WISP-3* differs from all CCN family members described previously, in that it contains only six of the 10 cysteine residues (Fig. 3A and B). A short variable region follows the VWC domain. The third module, the thrombospondin (TSP) domain is involved in binding to sulfated glycoconjugates and contains six cysteine residues and a conserved WSxCSxxCG motif first identified in thrombospondin (25). The C-terminal (CT) module containing the remaining 10 cysteines is thought to be involved in dimerization and receptor binding (26). The CT domain is present in all CCN family members described to date but is absent in *WISP-2* (Fig. 3A and B). The existence of a putative signal sequence and the absence of a transmembrane domain suggest that *WISPs* are secreted proteins, an observation supported by an analysis of their expression and secretion from mammalian cell and baculovirus cultures (data not shown).

**Expression of *WISP* mRNA in Human Tissues.** Tissue-specific expression of human *WISPs* was characterized by PCR

analysis on adult and fetal multiple tissue cDNA panels. *WISP-1* expression was seen in the adult heart, kidney, lung, pancreas, placenta, ovary, small intestine, and spleen (Fig. 3C). Little or no expression was detected in the brain, liver, skeletal muscle, colon, peripheral blood leukocytes, prostate, testis, or thymus. *WISP-2* had a more restricted tissue expression and was detected in adult skeletal muscle, colon, ovary, and fetal lung. Predominant expression of *WISP-3* was seen in adult kidney and testis and fetal kidney. Lower levels of *WISP-3* expression were detected in placenta, ovary, prostate, and small intestine.

**In Situ Localization of *WISP-1* and *WISP-2*.** Expression of *WISP-1* and *WISP-2* was assessed by *in situ* hybridization in mammary tumors from Wnt-1 transgenic mice. Strong expression of *WISP-1* was observed in stromal fibroblasts lying within the fibrovascular tumor stroma (Fig. 4 A–D). However, low-level *WISP-1* expression also was observed focally within tumor cells (data not shown). No expression was observed in normal breast. Like *WISP-1*, *WISP-2* expression also was seen in the tumor stroma in breast tumors from Wnt-1 transgenic animals (Fig. 4 E–H). However, *WISP-2* expression in the stroma was in spindle-shaped cells adjacent to capillary vessels, whereas



**FIG. 4.** (A, C, E, and G) Representative hematoxylin/eosin-stained images from breast tumors in Wnt-1 transgenic mice. The corresponding dark-field images showing *WISP-1* expression are shown in B and D. The tumor is a moderately well-differentiated adenocarcinoma showing evidence of adenoid cystic change. At low power (A and B), expression of *WISP-1* is seen in the delicate branching fibrovascular tumor stroma (arrowhead). At higher magnification, expression is seen in the stromal(s) fibroblasts (C and D), and tumor cells are negative. Focal expression of *WISP-1*, however, was observed in tumor cells in some areas. Images of *WISP-2* expression are shown in E–H. At low power (E and F), expression of *WISP-2* is seen in cells lying within the fibrovascular tumor stroma. At higher magnification, these cells appeared to be adjacent to capillary vessels whereas tumor cells are negative (G and H).

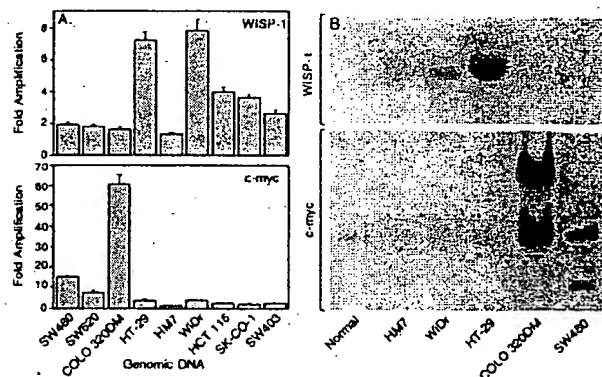
the predominant cell type expressing *WISP-1* was the stromal fibroblasts.

**Chromosome Localization of the *WISP* Genes.** The chromosomal location of the human *WISP* genes was determined by radiation hybrid mapping panels. *WISP-1* is approximately 3.48 cR from the meiotic marker AFM259xc5 [logarithm of odds (lod) score 16.31] on chromosome 8q24.1 to 8q24.3, in the same region as the human locus of the *novH* family member (27) and roughly 4 Mbs distal to *c-myc* (28). Preliminary fine mapping indicates that *WISP-1* is located near D8S1712 STS. *WISP-2* is linked to the marker SHGC-33922 (lod = 1,000) on chromosome 20q12–20q13.1. Human *WISP-3* mapped to chromosome 6q22–6q23 and is linked to the marker AFM211ze5 (lod = 1,000). *WISP-3* is approximately 18 Mbs proximal to CTGF and 23 Mbs proximal to the human cellular oncogene *MYB* (27, 29).

**Amplification and Aberrant Expression of *WISPs* in Human Colon Tumors.** Amplification of protooncogenes is seen in many human tumors and has etiological and prognostic significance. For example, in a variety of tumor types, *c-myc* amplification has been associated with malignant progression and poor prognosis (30). Because *WISP-1* resides in the same general chromosomal location (8q24) as *c-myc*, we asked whether it was a target of gene amplification, and, if so, whether this amplification was independent of the *c-myc* locus. Genomic DNA from human colon cancer cell lines was assessed by quantitative PCR and Southern blot analysis. (Fig. 5 A and B). Both methods detected similar degrees of *WISP-1* amplification. Most cell lines showed significant (2- to 4-fold) amplification, with the HT-29 and WiDr cell lines demonstrating an 8-fold increase. Significantly, the pattern of amplification observed did not correlate with that observed for *c-myc*, indicating that the *c-myc* gene is not part of the amplicon that involves the *WISP-1* locus.

We next examined whether the *WISP* genes were amplified in a panel of 25 primary human colon adenocarcinomas. The relative *WISP* gene copy number in each colon tumor DNA was compared with pooled normal DNA from 10 donors by quantitative PCR (Fig. 6). The copy number of *WISP-1* and *WISP-2* was significantly greater than one, approximately 2-fold for *WISP-1* in about 60% of the tumors and 2- to 4-fold for *WISP-2* in 92% of the tumors ( $P < 0.001$  for each). The copy number for *WISP-3* was indistinguishable from one ( $P = 0.166$ ). In addition, the copy number of *WISP-2* was significantly higher than that of *WISP-1* ( $P < 0.001$ ).

The levels of *WISP* transcripts in RNA isolated from 19 adenocarcinomas and their matched normal mucosa were



**FIG. 5.** Amplification of *WISP-1* genomic DNA in colon cancer cell lines. (A) Amplification in cell line DNA was determined by quantitative PCR. (B) Southern blots containing genomic DNA (10  $\mu$ g) digested with *EcoRI* (*WISP-1*) or *XbaI* (*c-myc*) were hybridized with a 100-bp human *WISP-1* probe (amino acids 186–219) or a human *c-myc* probe (located at bp 1901–2000). The *WISP* and *myc* genes are detected in normal human genomic DNA after a longer film exposure.

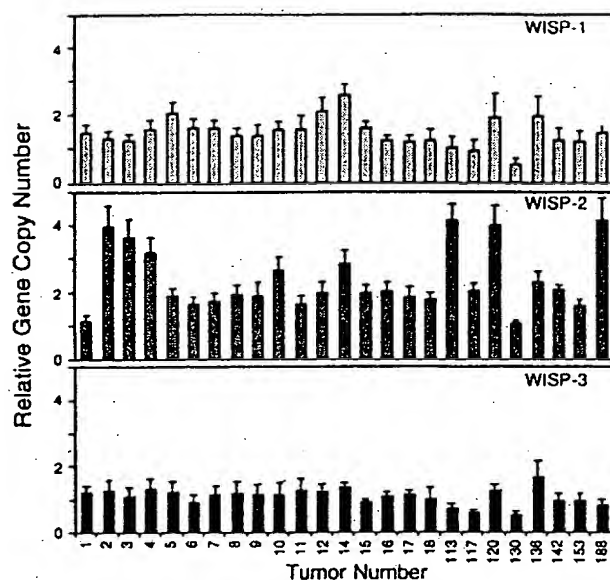


Fig. 6. Genomic amplification of *WISP* genes in human colon tumors. The relative gene copy number of the *WISP* genes in 25 adenocarcinomas was assayed by quantitative PCR, by comparing DNA from primary human tumors with pooled DNA from 10 healthy donors. The data are means  $\pm$  SEM from one experiment done in triplicate. The experiment was repeated at least three times.

assessed by quantitative PCR (Fig. 7). The level of *WISP-1* RNA present in tumor tissue varied but was significantly increased (2- to >25-fold) in 84% (16/19) of the human colon tumors examined compared with normal adjacent mucosa. Four of 19 tumors showed greater than 10-fold overexpression. In contrast, in 79% (15/19) of the tumors examined, *WISP-2* RNA expression was significantly lower in the tumor than the mucosa. Similar to *WISP-1*, *WISP-3* RNA was overexpressed in 63% (12/19) of the colon tumors compared with the normal

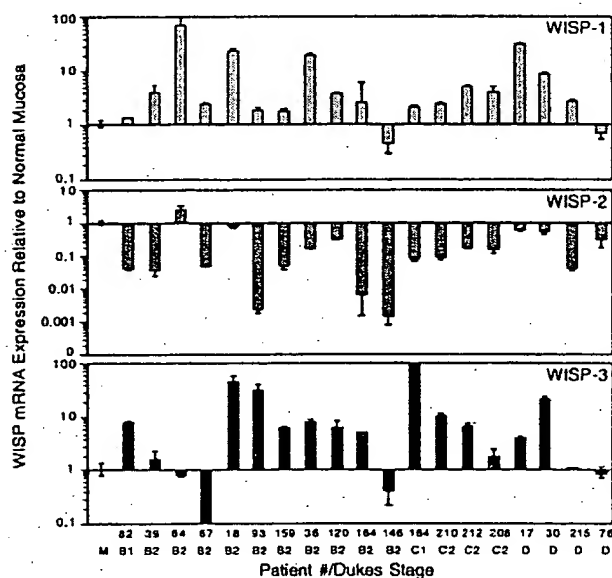


Fig. 7. *WISP* RNA expression in primary human colon tumors relative to expression in normal mucosa from the same patient. Expression of *WISP* mRNA in 19 adenocarcinomas was assayed by quantitative PCR. The Dukes stage of the tumor is listed under the sample number. The data are means  $\pm$  SEM from one experiment done in triplicate. The experiment was repeated at least twice.

mucosa. The amount of overexpression of *WISP-3* ranged from 4- to >40-fold.

## DISCUSSION

One approach to understanding the molecular basis of cancer is to identify differences in gene expression between cancer cells and normal cells. Strategies based on assumptions that steady-state mRNA levels will differ between normal and malignant cells have been used to clone differentially expressed genes (31). We have used a PCR-based selection strategy, SSH, to identify genes selectively expressed in C57MG mouse mammary epithelial cells, transformed by Wnt-1.

Three of the genes isolated, *WISP-1*, *WISP-2*, and *WISP-3*, are members of the CCN family of growth factors, which includes CTGF, Cyr61, and *nov*, a family not previously linked to Wnt signaling.

Two independent experimental systems demonstrated that *WISP* induction was associated with the expression of Wnt-1. The first was C57MG cells infected with a Wnt-1 retroviral vector or C57MG cells expressing Wnt-1 under the control of a tetracycline-repressible promoter, and the second was in Wnt-1 transgenic mice, where breast tissue expresses Wnt-1, whereas normal breast tissue does not. No *WISP* RNA expression was detected in mammary tumors induced by polyoma virus middle T antigen (data not shown). These data suggest a link between Wnt-1 and *WISPs* in that in these two situations, *WISP* induction was correlated with Wnt-1 expression.

It is not clear whether the *WISPs* are directly or indirectly induced by the downstream components of the Wnt-1 signaling pathway (i.e.,  $\beta$ -catenin-TCF-1/Lef1). The increased levels of *WISP* RNA were measured in Wnt-1-transformed cells, hours or days after Wnt-1 transformation. Thus, *WISP* expression could result from Wnt-1 signaling directly through  $\beta$ -catenin transcription factor regulation or alternatively through Wnt-1 signaling turning on a transcription factor, which in turn regulates *WISPs*.

The *WISPs* define an additional subfamily of the CCN family of growth factors. One striking difference observed in the protein sequence of *WISP-2* is the absence of a CT domain, which is present in CTGF, Cyr61, *nov*, *WISP-1*, and *WISP-3*. This domain is thought to be involved in receptor binding and dimerization. Growth factors, such as TGF- $\beta$ , platelet-derived growth factor, and nerve growth factor, which contain a cystine knot motif exist as dimers (32). It is tempting to speculate that *WISP-1* and *WISP-3* may exist as dimers, whereas *WISP-2* exists as a monomer. If the CT domain is also important for receptor binding, *WISP-2* may bind its receptor through a different region of the molecule than the other CCN family members. No specific receptors have been identified for CTGF or *nov*. A recent report has shown that integrin  $\alpha_v\beta_3$  serves as an adhesion receptor for Cyr61 (33).

The strong expression of *WISP-1* and *WISP-2* in cells lying within the fibrovascular tumor stroma in breast tumors from Wnt-1 transgenic animals is consistent with previous observations that transcripts for the related CTGF gene are primarily expressed in the fibrous stroma of mammary tumors (34). Epithelial cells are thought to control the proliferation of connective tissue stroma in mammary tumors by a cascade of growth factor signals similar to that controlling connective tissue formation during wound repair. It has been proposed that mammary tumor cells or inflammatory cells at the tumor interstitial interface secrete TGF- $\beta$ 1, which is the stimulus for stromal proliferation (34). TGF- $\beta$ 1 is secreted by a large percentage of malignant breast tumors and may be one of the growth factors that stimulates the production of CTGF and *WISPs* in the stroma.

It was of interest that *WISP-1* and *WISP-2* expression was observed in the stromal cells that surrounded the tumor cells

(epithelial cells) in the Wnt-1 transgenic mouse sections of breast tissue. This finding suggests that paracrine signaling could occur in which the stromal cells could supply WISP-1 and WISP-2 to regulate tumor cell growth on the WISP extracellular matrix. Stromal cell-derived factors in the extracellular matrix have been postulated to play a role in tumor cell migration and proliferation (35). The localization of *WISP-1* and *WISP-2* in the stromal cells of breast tumors supports this paracrine model.

An analysis of *WISP-1* gene amplification and expression in human colon tumors showed a correlation between DNA amplification and overexpression, whereas overexpression of *WISP-3* RNA was seen in the absence of DNA amplification. In contrast, *WISP-2* DNA was amplified in the colon tumors, but its mRNA expression was significantly reduced in the majority of tumors compared with the expression in normal colonic mucosa from the same patient. The gene for human *WISP-2* was localized to chromosome 20q12–20q13, at a region frequently amplified and associated with poor prognosis in node negative breast cancer and many colon cancers, suggesting the existence of one or more oncogenes at this locus (36–38). Because the center of the 20q13 amplicon has not yet been identified, it is possible that the apparent amplification observed for *WISP-2* may be caused by another gene in this amplicon.

A recent manuscript on *rCop-1*, the rat orthologue of *WISP-2*, describes the loss of expression of this gene after cell transformation, suggesting it may be a negative regulator of growth in cell lines (16). Although the mechanism by which *WISP-2* RNA expression is down-regulated during malignant transformation is unknown, the reduced expression of *WISP-2* in colon tumors and cell lines suggests that it may function as a tumor suppressor. These results show that the *WISP* genes are aberrantly expressed in colon cancer and suggest that their altered expression may confer selective growth advantage to the tumor.

Members of the Wnt signaling pathway have been implicated in the pathogenesis of colon cancer, breast cancer, and melanoma, including the tumor suppressor gene adenomatous polyposis coli and  $\beta$ -catenin (39). Mutations in specific regions of either gene can cause the stabilization and accumulation of cytoplasmic  $\beta$ -catenin, which presumably contributes to human carcinogenesis through the activation of target genes such as the *WISPs*. Although the mechanism by which Wnt-1 transforms cells and induces tumorigenesis is unknown, the identification of *WISPs* as genes that may be regulated downstream of Wnt-1 in C57MG cells suggests they could be important mediators of Wnt-1 transformation. The amplification and altered expression patterns of the *WISPs* in human colon tumors may indicate an important role for these genes in tumor development.

We thank the DNA synthesis group for oligonucleotide synthesis, T. Baker for technical assistance, P. Dowd for radiation hybrid mapping, K. Willert and R. Nusse for the tet-repressible C57MG/Wnt-1 cells, V. Dixit for discussions, and D. Wood and A. Bruce for artwork.

- Cadigan, K. M. & Nusse, R. (1997) *Genes Dev.* 11, 3286–3305.
- Dale, T. C. (1998) *Biochem. J.* 329, 209–223.
- Nusse, R. & Varmus, H. E. (1982) *Cell* 31, 99–109.
- van Ooyen, A. & Nusse, R. (1984) *Cell* 39, 233–240.
- Tsukamoto, A. S., Grosschedl, R., Guzman, R. C., Parslow, T. & Varmus, H. E. (1988) *Cell* 55, 619–625.
- Brown, J. D. & Moon, R. T. (1998) *Curr. Opin. Cell Biol.* 10, 182–187.
- Molenaar, M., van de Wetering, M., Oosterwegel, M., Peterson-Maduro, J., Godsave, S., Korinek, V., Roose, J., Destree, O. & Clevers, H. (1996) *Cell* 86, 391–399.
- Korinek, V., Barker, N., Willert, K., Molenaar, M., Roose, J., Wagenaar, G., Markman, M., Lamers, W., Destree, O. & Clevers, H. (1998) *Mol. Cell Biol.* 18, 1248–1256.
- Munemitsu, S., Albert, I., Souza, B., Rubinfeld, B. & Polakis, P. (1995) *Proc. Natl. Acad. Sci. USA* 92, 3046–3050.
- He, T. C., Sparks, A. B., Rago, C., Hermeking, H., Zawel, L., da Costa, L. T., Morin, P. J., Vogelstein, B. & Kinzler, K. W. (1998) *Science* 281, 1509–1512.
- Diatchenko, L., Lau, Y. F., Campbell, A. P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E. D. & Siebert, P. D. (1996) *Proc. Natl. Acad. Sci. USA* 93, 6025–6030.
- Brown, A. M., Wildin, R. S., Prendergast, T. J. & Varmus, H. E. (1986) *Cell* 46, 1001–1009.
- Wong, G. T., Gavin, B. J. & McMahon, A. P. (1994) *Mol. Cell Biol.* 14, 6278–6286.
- Shimizu, H., Julius, M. A., Giarre, M., Zheng, Z., Brown, A. M. & Kitajewski, J. (1997) *Cell Growth Differ.* 8, 1349–1358.
- Hashimoto, Y., Shindo-Okada, N., Tani, M., Nagamachi, Y., Takeuchi, K., Shiroishi, T., Toma, H. & Yokota, J. (1998) *J. Exp. Med.* 187, 289–296.
- Zhang, R., Averboukh, L., Zhu, W., Zhang, H., Jo, H., Dempsey, P. J., Coffey, R. J., Pardee, A. B. & Liang, P. (1998) *Mol. Cell Biol.* 18, 6131–6141.
- Grotendorst, G. R. (1997) *Cytokine Growth Factor Rev.* 8, 171–179.
- Kireeva, M. L., Mo, F. E., Yang, G. P. & Lau, L. F. (1996) *Mol. Cell Biol.* 16, 1326–1334.
- Babic, A. M., Kireeva, M. L., Kolesnikova, T. V. & Lau, L. F. (1998) *Proc. Natl. Acad. Sci. USA* 95, 6355–6360.
- Martinerie, C., Huff, V., Joubert, I., Badzioch, M., Saunders, G., Strong, L. & Perbal, B. (1994) *Oncogene* 9, 2729–2732.
- Bork, P. (1993) *FEBS Lett.* 327, 125–130.
- Kim, H. S., Nagalla, S. R., Oh, Y., Wilson, E., Roberts, C. T., Jr. & Rosenfeld, R. G. (1997) *Proc. Natl. Acad. Sci. USA* 94, 12981–12986.
- Joliet, V., Martinerie, C., Dambrine, G., Plassiart, G., Brisac, M., Crochet, J. & Perbal, B. (1992) *Mol. Cell Biol.* 12, 10–21.
- Mancuso, D. J., Tuley, E. A., Westfield, L. A., Worrall, N. K., Shelton-Inloes, B. B., Sorace, J. M., Alevy, Y. G. & Sadler, J. E. (1989) *J. Biol. Chem.* 264, 19514–19527.
- Holt, G. D., Pangburn, M. K. & Ginsburg, V. (1990) *J. Biol. Chem.* 265, 2852–2855.
- Voorberg, J., Fontijn, R., Calafat, J., Janssen, H., van Mourik, J. A. & Pannekoek, H. (1991) *J. Cell Biol.* 113, 195–205.
- Martinerie, C., Viegas-Pequignot, E., Guenard, I., Dutrillaux, B., Nguyen, V. C., Bernheim, A. & Perbal, B. (1992) *Oncogene* 7, 2529–2534.
- Takahashi, E., Hori, T., O'Connell, P., Leppert, M. & White, R. (1991) *Cytogenet. Cell Genet.* 57, 109–111.
- Meesse, E., Meltzer, P. S., Witkowski, C. M. & Trent, J. M. (1989) *Genes Chromosomes Cancer* 1, 88–94.
- Garte, S. J. (1993) *Crit. Rev. Oncog.* 4, 435–449.
- Zhang, L., Zhou, W., Velculescu, V. E., Kern, S. E., Hruban, R. H., Hamilton, S. R., Vogelstein, B. & Kinzler, K. W. (1997) *Science* 276, 1268–1272.
- Sun, P. D. & Davies, D. R. (1995) *Annu. Rev. Biophys. Biomol. Struct.* 24, 269–291.
- Kireeva, M. L., Lam, S. C. T. & Lau, L. F. (1998) *J. Biol. Chem.* 273, 3090–3096.
- Frazier, K. S. & Grotendorst, G. R. (1997) *Int. J. Biochem. Cell Biol.* 29, 153–161.
- Wernert, N. (1997) *Virchows Arch.* 430, 433–443.
- Tanner, M. M., Tirkkonen, M., Kallioniemi, A., Collins, C., Stokke, T., Karhu, R., Kowbel, D., Shadravan, F., Hintz, M., Kuo, W. L., *et al.* (1994) *Cancer Res.* 54, 4257–4260.
- Brinkmann, U., Gallo, M., Polymeropoulos, M. H. & Pastan, I. (1996) *Genome Res.* 6, 187–194.
- Bischoff, J. R., Anderson, L., Zhu, Y., Mossie, K., Ng, L., Souza, B., Schryver, B., Flanagan, P., Clairvoyant, F., Ginther, C., *et al.* (1998) *EMBO J.* 17, 3052–3065.
- Morin, P. J., Sparks, A. B., Korinek, V., Barker, N., Clevers, H., Vogelstein, B. & Kinzler, K. W. (1997) *Science* 275, 1787–1790.
- Lu, L. H. & Gillett, N. (1994) *Cell Vision* 1, 169–176.



methods. Peptides AENK or AEQK were dissolved in water, made isotonic with NaCl and diluted into RPMI growth medium. T-cell-proliferation assays were done essentially as described<sup>20,21</sup>. Briefly, after antigen pulsing ( $30 \mu\text{g ml}^{-1}$  TTCF) with tetrapeptides ( $1-2 \text{ mg ml}^{-1}$ ), PBMCs or EBV-B cells were washed in PBS and fixed for 45 s in 0.05% glutaraldehyde. Glycine was added to a final concentration of 0.1M and the cells were washed five times in RPMI 1640 medium containing 1% FCS before co-culture with T-cell clones in round-bottom 96-well microtitre plates. After 48 h, the cultures were pulsed with  $1 \mu\text{Ci}$  of  $^3\text{H}$ -thymidine and harvested for scintillation counting 16 h later. Predigestion of native TTCF was done by incubating  $200 \mu\text{g}$  TTCF with  $0.25 \mu\text{g}$  pig kidney legumain in  $500 \mu\text{l}$  50 mM citrate buffer, pH 5.5, for 1 h at  $37^\circ\text{C}$ . **Glycopeptide digestions.** The peptides HIDNEEDI, HIDN(N-glucosamine) EEDI and HIDNESDI, which are based on the TTCF sequence, and QQQLHFGSNVTDCSGNFCLFR(KKK), which is based on human transferrin, were obtained by custom synthesis. The three C-terminal lysine residues were added to the natural sequence to aid solubility. The transferrin glycopeptide QQQLHFGSNVTDCSGNFCLFR was prepared by tryptic (Promega) digestion of 5 mg reduced, carboxy-methylated human transferrin followed by concanavalin A chromatography<sup>11</sup>. Glycopeptides corresponding to residues 622-642 and 421-452 were isolated by reverse-phase HPLC and identified by mass spectrometry and N-terminal sequencing. The lyophilized transferrin-derived peptides were redissolved in 50 mM sodium acetate, pH 5.5, 10 mM dithiothreitol, 20% methanol. Digestions were performed for 3 h at  $30^\circ\text{C}$  with  $5-50 \text{ mU ml}^{-1}$  pig kidney legumain or B-cell AEP. Products were analysed by HPLC or MALDI-TOF mass spectrometry using a matrix of  $10 \text{ mg ml}^{-1}$   $\alpha$ -cyanocinnamic acid in 50% acetonitrile/0.1% TFA and a PerSeptive Biosystems Elite STR mass spectrometer set to linear or reflector mode. Internal standardization was obtained with a matrix ion of 568.13 mass units.

Received 29 September; accepted 3 November 1998.

- Chen, J. M. et al. Cloning, isolation, and characterisation of mammalian legumain, an asparaginyl endopeptidase. *J. Biol. Chem.* 272, 8090-8098 (1997).
- Kembhavi, A. A., Butt, D. J., Knight, C. G. & Barrett, A. J. The two cysteine endopeptidases of legume seeds: purification and characterization by use of specific fluorometric assays. *Arch. Biochem. Biophys.* 303, 208-213 (1993).
- Dalton, J. P., Hla Jamsriska, L. & Bridley, P. J. Asparaginyl endopeptidase activity in adult *Schistosoma mansoni*. *Parasitology* 111, 575-580 (1995).
- Bennett, K. et al. Antigen processing for presentation by class II major histocompatibility complex requires cleavage by cathepsin E. *Eur. J. Immunol.* 22, 1519-1524 (1992).
- Riese, R. J. et al. Essential role for cathepsin S in MHC class II-associated invariant chain processing and peptide loading. *Immunity* 4, 357-366 (1996).
- Rodriguez, G. M. & Diment, S. Role of cathepsin D in antigen presentation of ovalbumin. *J. Immunol.* 149, 2894-2898 (1992).
- Hewitt, E. W. et al. Natural processing sites for human cathepsin E and cathepsin D in tetanus toxin: implications for T cell epitope generation. *J. Immunol.* 159, 4693-4699 (1997).
- Watts, C. Capture and processing of exogenous antigens for presentation on MHC molecules. *Annu. Rev. Immunol.* 15, 821-850 (1997).
- Chapman, H. A. Endosomal proteases and MHC class II function. *Curr. Opin. Immunol.* 10, 93-102 (1998).
- Fineschi, B. & Miller, J. Endosomal proteases and antigen processing. *Trends Biochem. Sci.* 22, 377-382 (1997).
- Lu, J. & van Halbeek, H. Complete  $^1\text{H}$  and  $^{13}\text{C}$  resonance assignments of a 21-amino acid glycopeptide prepared from human serum transferrin. *Carbohydr. Res.* 296, 1-21 (1996).
- Fearon, D. T. & Locksley, R. M. The instructive role of innate immunity in the acquired immune response. *Science* 272, 50-54 (1996).
- Medzhitov, R. & Janeway, C. A. J. Innate immunity: the virtues of a nonclonal system of recognition. *Cell* 91, 295-298 (1997).
- Wyatt, R. et al. The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature* 393, 705-711 (1998).
- Botarelli, P. et al. N-glycosylation of HIV gp120 may constrain recognition by T lymphocytes. *J. Immunol.* 147, 3128-3132 (1991).
- Davidson, H. W., West, M. A. & Watts, C. Endocytosis, intracellular trafficking, and processing of membrane IgG and monovalent antigen/membrane IgG complexes in B lymphocytes. *J. Immunol.* 144, 4101-4109 (1990).
- Barrett, A. J. & Kirschke, H. Cathepsin B, cathepsin H and cathepsin L. *Methods Enzymol.* 80, 535-559 (1981).
- Makoff, A. J., Ballantine, S. P., Smallwood, A. E. & Fairweather, N. F. Expression of tetanus toxin fragment C in *E. coli*: its purification and potential use as a vaccine. *Biotechnology* 7, 1043-1046 (1989).
- Lane, D. P. & Harlow, E. *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, 1988).
- Lanzavecchia, A. Antigen-specific interaction between T and B cells. *Nature* 314, 537-539 (1985).
- Pond, L. & Watts, C. Characterization of transport of newly assembled, T cell-stimulatory MHC class II-peptide complexes from MHC class II compartments to the cell surface. *J. Immunol.* 159, 543-553 (1997).

**Acknowledgements.** We thank M. Ferguson for helpful discussions and advice; E. Smythe and L. Grayson for advice and technical assistance; B. Spruce, A. Knight and the BTS (Ninewells Hospital) for help with blood monocyte preparation; and our colleagues for many helpful comments on the manuscript. This work was supported by the Wellcome Trust and by an EMBO Long-term fellowship to B. M.

Correspondence and requests for materials should be addressed to C.W. (e-mail: c.watts@dundee.ac.uk).

## Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer

Robert M. Pitti<sup>††</sup>, Scot A. Marsters<sup>††</sup>, David A. Lawrence<sup>††</sup>, Margaret Roy<sup>\*</sup>, Frank C. Kischkel<sup>\*</sup>, Patrick Dowd<sup>\*</sup>, Arthur Huang<sup>\*</sup>, Christopher J. Donahue<sup>\*</sup>, Steven W. Sherwood<sup>\*</sup>, Daryl T. Baldwin<sup>\*</sup>, Paul J. Godowski<sup>\*</sup>, William I. Wood<sup>\*</sup>, Austin L. Gurney<sup>\*</sup>, Kenneth J. Hillan<sup>\*</sup>, Robert L. Cohen<sup>\*</sup>, Audrey D. Goddard<sup>\*</sup>, David Botstein<sup>†</sup> & Avi Ashkenazi<sup>\*</sup>

<sup>\*</sup> Departments of Molecular Oncology, Molecular Biology, and Immunology, Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA

<sup>†</sup> Department of Genetics, Stanford University, Stanford, California 94305, USA

<sup>††</sup> These authors contributed equally to this work

Fas ligand (FasL) is produced by activated T cells and natural killer cells and it induces apoptosis (programmed cell death) in target cells through the death receptor Fas/Apo1/CD95 (ref. 1). One important role of FasL and Fas is to mediate immune-cytotoxic killing of cells that are potentially harmful to the organism, such as virus-infected or tumour cells<sup>1</sup>. Here we report the discovery of a soluble decoy receptor, termed decoy receptor 3 (Dcr3), that binds to FasL and inhibits FasL-induced apoptosis. The Dcr3 gene was amplified in about half of 35 primary lung and colon tumours studied, and Dcr3 messenger RNA was expressed in malignant tissue. Thus, certain tumours may escape FasL-dependent immune-cytotoxic attack by expressing a decoy receptor that blocks FasL.

By searching expressed sequence tag (EST) databases, we identified a set of related ESTs that showed homology to the tumour necrosis factor (TNF) receptor (TNFR) gene superfamily<sup>2</sup>. Using the overlapping sequence, we isolated a previously unknown full-length complementary DNA from human fetal lung. We named the protein encoded by this cDNA decoy receptor 3 (Dcr3). The cDNA encodes a 300-amino-acid polypeptide that resembles members of the TNFR family (Fig. 1a): the amino terminus contains a leader sequence, which is followed by four tandem cysteine-rich domains (CRDs). Like one other TNFR homologue, osteoprotegerin (OPG)<sup>3</sup>, Dcr3 lacks an apparent transmembrane sequence, which indicates that it may be a secreted, rather than a membrane-associated, molecule. We expressed a recombinant, histidine-tagged form of Dcr3 in mammalian cells; Dcr3 was secreted into the cell culture medium, and migrated on polyacrylamide gels as a protein of relative molecular mass 35,000 (data not shown). Dcr3 shares sequence identity in particular with OPG (31%) and TNFR2 (29%), and has relatively less homology with Fas (17%). All of the cysteines in the four CRDs of Dcr3 and OPG are conserved; however, the carboxy-terminal portion of Dcr3 is 101 residues shorter.

We analysed expression of Dcr3 mRNA in human tissues by northern blotting (Fig. 1b). We detected a predominant 1.2-kilobase transcript in fetal lung, brain, and liver, and in adult spleen, colon and lung. In addition, we observed relatively high Dcr3 mRNA expression in the human colon carcinoma cell line SW480.

To investigate potential ligand interactions of Dcr3, we generated a recombinant, Fc-tagged Dcr3 protein. We tested binding of Dcr3-Fc to human 293 cells transfected with individual TNF-family ligands, which are expressed as type 2 transmembrane proteins (these transmembrane proteins have their N termini in the cytosol). Dcr3-Fc showed a significant increase in binding to cells transfected with FasL<sup>4</sup> (Fig. 2a), but not to cells transfected with TNF<sup>5</sup>, Apo2L/TRAIL<sup>6,7</sup>, Apo3L/TWEAK<sup>8,9</sup>, or OPGL/TRANCE/

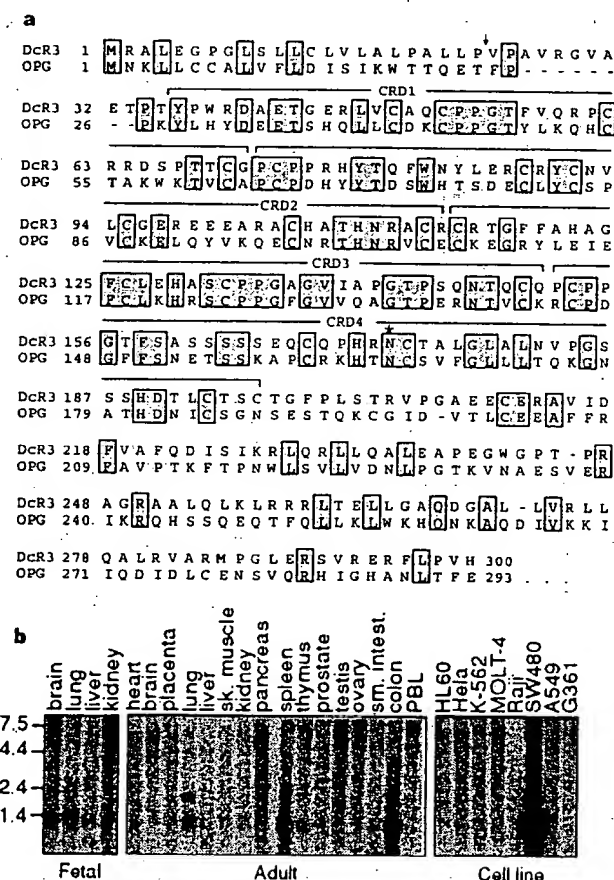
RANKL<sup>10-12</sup> (data not shown). DcR3-Fc immunoprecipitated shed FasL from FasL-transfected 293 cells (Fig. 2b) and purified soluble FasL (Fig. 2c), as did the Fc-tagged ectodomain of Fas but not TNFR1. Gel-filtration chromatography showed that DcR3-Fc and soluble FasL formed a stable complex (Fig. 2d). Equilibrium analysis indicated that DcR3-Fc and Fas-Fc bound to soluble FasL with a comparable affinity ( $K_d = 0.8 \pm 0.2$  and  $1.1 \pm 0.1$  nM, respectively; Fig. 2e), and that DcR3-Fc could block nearly all of the binding of soluble FasL to Fas-Fc (Fig. 2e, inset). Thus, DcR3 competes with Fas for binding to FasL.

To determine whether binding of DcR3 inhibits FasL activity, we tested the effect of DcR3-Fc on apoptosis induction by soluble FasL in Jurkat T leukaemia cells, which express Fas (Fig. 3a). DcR3-Fc and Fas-Fc blocked soluble-FasL-induced apoptosis in a similar dose-dependent manner, with half-maximal inhibition at  $\sim 0.1 \mu\text{g ml}^{-1}$ . Time-course analysis showed that the inhibition did not merely delay cell death, but rather persisted for at least 24 hours (Fig. 3b). We also tested the effect of DcR3-Fc on activation-induced cell death (AICD) of mature T lymphocytes, a FasL-dependent process<sup>1</sup>. Consistent with previous results<sup>13</sup>, activation of interleukin-2-stimulated CD4-positive T cells with anti-CD3 antibody increased the level of apoptosis twofold, and Fas-Fc blocked this effect substantially (Fig. 3c); DcR3-Fc blocked the

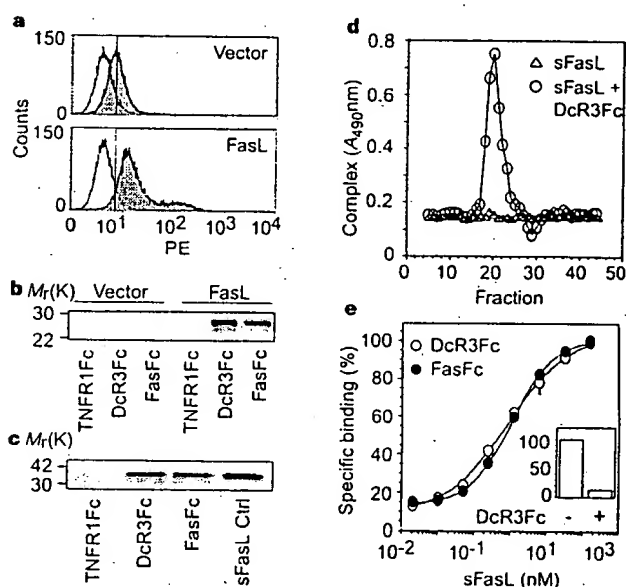
induction of apoptosis to a similar extent. Thus, DcR3 binding blocks apoptosis induction by FasL.

FasL-induced apoptosis is important in elimination of virus-infected cells and cancer cells by natural killer cells and cytotoxic T lymphocytes; an alternative mechanism involves perforin and granzymes<sup>14-16</sup>. Peripheral blood natural killer cells triggered marked cell death in Jurkat T leukaemia cells (Fig. 3d); DcR3-Fc and Fas-Fc each reduced killing of target cells from  $\sim 65\%$  to  $\sim 30\%$ , with half-maximal inhibition at  $\sim 1 \mu\text{g ml}^{-1}$ ; the residual killing was probably mediated by the perforin/granzyme pathway. Thus, DcR3 binding blocks FasL-dependent natural killer cell activity. Higher DcR3-Fc and Fas-Fc concentrations were required to block natural killer cell activity compared with those required to block soluble FasL activity, which is consistent with the greater potency of membrane-associated FasL compared with soluble FasL<sup>17</sup>.

Given the role of immune-cytotoxic cells in elimination of tumour cells and the fact that DcR3 can act as an inhibitor of FasL, we proposed that DcR3 expression might contribute to the ability of some tumours to escape immune-cytotoxic attack. As genomic amplification frequently contributes to tumorigenesis, we investigated whether the DcR3 gene is amplified in cancer. We analysed DcR3 gene-copy number by quantitative polymerase chain



**Figure 1** Primary structure and expression of human DcR3. **a**, Alignment of the amino-acid sequences of DcR3 and of osteoprotegerin (OPG); the C-terminal 101 residues of OPG are not shown. The putative signal cleavage site (arrow), the cysteine-rich domains (CRD 1-4), and the N-linked glycosylation site (asterisk) are shown. **b**, Expression of DcR3 mRNA. Northern hybridization analysis was done using the DcR3 cDNA as a probe and blots of poly(A)<sup>+</sup> RNA (Clontech) from human fetal and adult tissues or cancer cell lines. PBL, peripheral blood lymphocyte.



**Figure 2** Interaction of DcR3 with FasL. **a**, 293 cells were transfected with pRK5 vector (top) or with pRK5 encoding full-length FasL (bottom), incubated with DcR3-Fc (solid line, shaded area), TNFR1-Fc (dotted line) or buffer control (dashed line) (the dashed and dotted lines overlap), and analysed for binding by FACS. Statistical analysis showed a significant difference ( $P < 0.001$ ) between the binding of DcR3-Fc to cells transfected with FasL or pRK5. PE, phycoerythrin-labelled cells. **b**, 293 cells were transfected as in **a** and metabolically labelled, and cell supernatants were immunoprecipitated with Fc-tagged TNFR1, DcR3 or Fas. **c**, Purified soluble FasL (sFasL) was immunoprecipitated with TNFR1-Fc, DcR3-Fc or Fas-Fc and visualized by immunoblot with anti-FasL antibody. sFasL was loaded directly for comparison in the right-hand lane. **d**, Flag-tagged sFasL was incubated with DcR3-Fc or with buffer and resolved by gel filtration; column fractions were analysed in an assay that detects complexes containing DcR3-Fc and sFasL-Flag. **e**, Equilibrium binding of DcR3-Fc or Fas-Fc to sFasL-Flag. Inset, competition of DcR3-Fc with Fas-Fc for binding to sFasL-Flag.

reaction (PCR)<sup>18</sup> in genomic DNA from 35 primary lung and colon tumours, relative to pooled genomic DNA from peripheral blood leukocytes (PBLs) of 10 healthy donors. Eight of 18 lung tumours and 9 of 17 colon tumours showed DcR3 gene amplification, ranging from 2- to 18-fold (Fig. 4a, b). To confirm this result, we analysed the colon tumour DNAs with three more, independent sets of DcR3-based PCR primers and probes; we observed nearly the same amplification (data not shown).

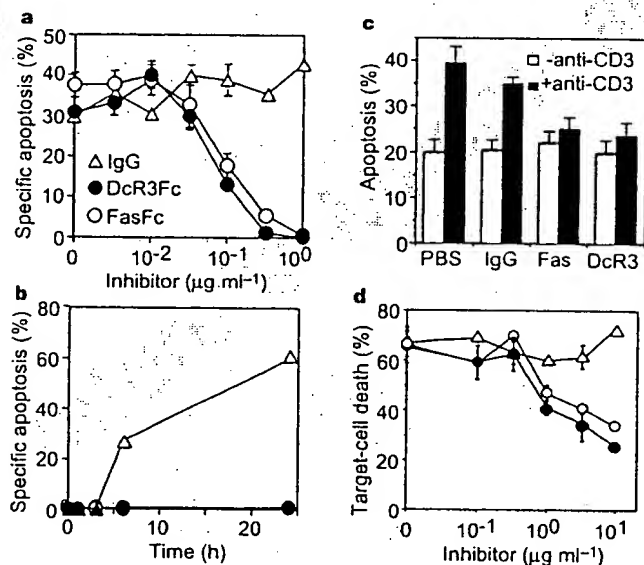
We then analysed DcR3 mRNA expression in primary tumour tissue sections by *in situ* hybridization. We detected DcR3 expression in 6 out of 15 lung tumours, 2 out of 2 colon tumours, 2 out of 5 breast tumours, and 1 out of 1 gastric tumour (data not shown). A section through a squamous-cell carcinoma of the lung is shown in Fig. 4c. DcR3 mRNA was localized to infiltrating malignant epithelium, but was essentially absent from adjacent stroma, indicating tumour-specific expression. Although the individual tumour specimens that we analysed for mRNA expression and gene amplification were different, the *in situ* hybridization results are consistent with the finding that the DcR3 gene is amplified frequently in tumours. SW480 colon carcinoma cells, which showed abundant DcR3 mRNA expression (Fig. 1b), also had marked DcR3 gene amplification, as shown by quantitative PCR (fourfold) and by Southern blot hybridization (fivefold) (data not shown).

If DcR3 amplification in cancer is functionally relevant, then DcR3 should be amplified more than neighbouring genomic regions that are not important for tumour survival. To test this,

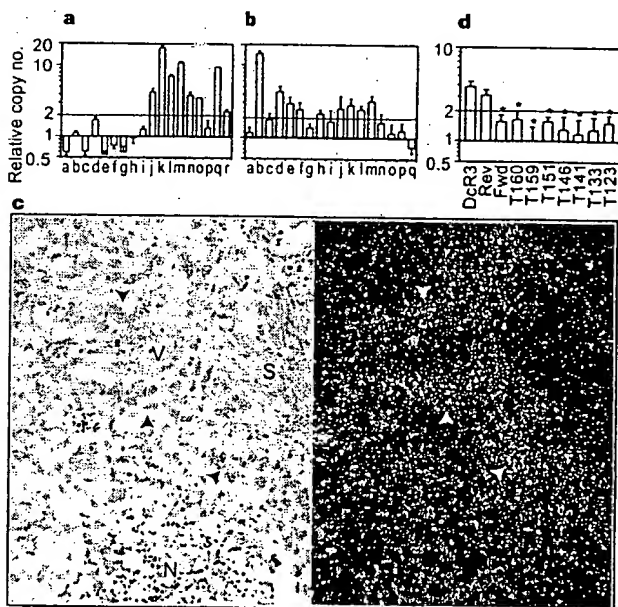
we mapped the human DcR3 gene by radiation-hybrid analysis; DcR3 showed linkage to marker AFM218xe7 (T160), which maps to chromosome position 20q13. Next, we isolated from a bacterial artificial chromosome (BAC) library a human genomic clone that carries DcR3, and sequenced the ends of the clone's insert. We then determined, from the nine colon tumours that showed twofold or greater amplification of DcR3, the copy number of the DcR3-flanking sequences (reverse and forward) from the BAC, and of seven genomic markers that span chromosome 20 (Fig. 4d). The DcR3-linked reverse marker showed an average amplification of roughly threefold, slightly less than the approximately fourfold amplification of DcR3; the other markers showed little or no amplification. These data indicate that DcR3 may be at the 'epicentre' of a distal chromosome 20 region that is amplified in colon cancer, consistent with the possibility that DcR3 amplification promotes tumour survival.

Our results show that DcR3 binds specifically to FasL and inhibits FasL activity. We did not detect DcR3 binding to several other TNF-ligand-family members; however, this does not rule out the possibility that DcR3 interacts with other ligands, as do some other TNFR family members, including OPG<sup>2,19</sup>.

FasL is important in regulating the immune response; however, little is known about how FasL function is controlled. One mechanism involves the molecule cFLIP, which modulates apoptosis signalling downstream of Fas<sup>20</sup>. A second mechanism involves proteolytic shedding of FasL from the cell surface<sup>17</sup>. DcR3 competes with Fas for



**Figure 3** Inhibition of FasL activity by DcR3. **a**, Human Jurkat T leukaemia cells were incubated with Flag-tagged soluble FasL (sFasL; 5 ng ml<sup>-1</sup>) oligomerized with anti-Flag antibody (0.1 μg ml<sup>-1</sup>) in the presence of the proposed inhibitors DcR3-Fc, Fas-Fc or human IgG1 and assayed for apoptosis (mean ± s.e.m. of triplicates). **b**, Jurkat cells were incubated with sFasL-Flag plus anti-Flag antibody as in **a**, in presence of 1 μg ml<sup>-1</sup> DcR3-Fc (filled circles), Fas-Fc (open circles) or human IgG1 (triangles), and apoptosis was determined at the indicated time points. **c**, Peripheral blood T cells were stimulated with PHA and interleukin-2, followed by control (white bars) or anti-CD3 antibody (filled bars), together with phosphate-buffered saline (PBS), human IgG1, Fas-Fc, or DcR3-Fc (10 μg ml<sup>-1</sup>). After 16 h, apoptosis of CD4<sup>+</sup> cells was determined (mean ± s.e.m. of results from five donors). **d**, Peripheral blood natural killer cells were incubated with <sup>51</sup>Cr-labelled Jurkat cells in the presence of DcR3-Fc (filled circles), Fas-Fc (open circles) or human IgG1 (triangles), and target-cell death was determined by release of <sup>51</sup>Cr (mean ± s.d. for two donors, each in triplicate).



**Figure 4** Genomic amplification of DcR3 in tumours. **a**, Lung cancers, comprising eight adenocarcinomas (c, d, f, g, h, j, k, l), seven squamous-cell carcinomas (a, e, m, n, o, p, q), one non-small-cell carcinoma (b), one small-cell carcinoma (i), and one bronchial adenocarcinoma (l). The data are means ± s.d. of 2 experiments done in duplicate. **b**, Colon tumours, comprising 17 adenocarcinomas. Data are means ± s.e.m. of five experiments done in duplicate. **c**, *In situ* hybridization analysis of DcR3 mRNA expression in a squamous-cell carcinoma of the lung. A representative bright-field image (left) and the corresponding dark-field image (right) show DcR3 mRNA over infiltrating malignant epithelium (arrowheads). Adjacent non-malignant stroma (S), blood vessel (V) and necrotic tumour tissue (N) are also shown. **d**, Average amplification of DcR3 compared with amplification of neighbouring genomic regions (reverse and forward, Rev and Fwd), the DcR3-linked marker T160, and other chromosome-20 markers, in the nine colon tumours showing DcR3 amplification of twofold or more (**b**). Data are from two experiments done in duplicate. Asterisk indicates  $P < 0.01$  for a Student's *t*-test comparing each marker with DcR3.



FasL binding; hence, it may represent a third mechanism of extracellular regulation of FasL activity. A decoy receptor that modulates the function of the cytokine interleukin-1 has been described<sup>21</sup>. In addition, two decoy receptors that belong to the TNFR family, DcR1 and DcR2, regulate the FasL-related apoptosis-inducing molecule Apo2L<sup>22</sup>. Unlike DcR1 and DcR2, which are membrane-associated proteins, DcR3 is directly secreted into the extracellular space. One other secreted TNFR-family member is OPG<sup>3</sup>, which shares greater sequence homology with DcR3 (31%) than do DcR1 (17%) or DcR2 (19%); OPG functions as a third decoy for Apo2L<sup>19</sup>. Thus, DcR3 and OPG define a new subset of TNFR-family members that function as secreted decoys to modulate ligands that induce apoptosis. Pox viruses produce soluble TNFR homologues that neutralize specific TNF-family ligands, thereby modulating the antiviral immune response<sup>2</sup>. Our results indicate that a similar mechanism, namely, production of a soluble decoy receptor for FasL, may contribute to immune evasion by certain tumours. □

## Methods

**Isolation of DcR3 cDNA.** Several overlapping ESTs in GenBank (accession numbers AA025672, AA025673 and W67560) and in Lifeseq<sup>TM</sup> (Incyte Pharmaceuticals; accession numbers 1339238, 1533571, 1533650, 1542861, 1789372 and 2207027) showed similarity to members of the TNFR family. We screened human cDNA libraries by PCR with primers based on the region of EST consensus; fetal lung was positive for a product of the expected size. By hybridization to a PCR-generated probe based on the ESTs, one positive clone (DNA30942) was identified. When searching for potential alternatively spliced forms of DcR3 that might encode a transmembrane protein, we isolated 50 more clones; the coding regions of these clones were identical in size to that of the initial clone (data not shown).

**Fc-fusion proteins (immunoadhesins).** The entire DcR3 sequence, or the ectodomain of Fas or TNFR1, was fused to the hinge and Fc region of human IgG1, expressed in insect SF9 cells or in human 293 cells, and purified as described<sup>23</sup>.

**Fluorescence-activated cell sorting (FACS) analysis.** We transfected 293 cells using calcium phosphate or Effectene (Qiagen) with pRK5 vector or pRK5 encoding full-length human FasL<sup>4</sup> (2 µg), together with pRK5 encoding CrmA (2 µg) to prevent cell death. After 16 h, the cells were incubated with biotinylated DcR3-Fc or TNFR1-Fc and then with phycoerythrin-conjugated streptavidin (GibcoBRL); and were assayed by FACS. The data were analysed by Kolmogorov-Smirnov statistical analysis. There was some detectable staining of vector-transfected cells by DcR3-Fc; as these cells express little FasL (data not shown), it is possible that DcR3 recognized some other factor that is expressed constitutively on 293 cells.

**Immunoprecipitation.** Human 293 cells were transfected as above, and metabolically labelled with [<sup>35</sup>S]cysteine and [<sup>35</sup>S]methionine (0.5 mCi; Amersham). After 16 h of culture in the presence of z-VAD-fmk (10 µM), the medium was immunoprecipitated with DcR3-Fc, Fas-Fc or TNFR1-Fc (5 µg), followed by protein A-Sepharose (Repligen). The precipitates were resolved by SDS-PAGE and visualized on a phosphorimager (Fuji BAS2000). Alternatively, purified, Flag-tagged soluble FasL (1 µg) (Alexis) was incubated with each Fc-fusion protein (1 µg), precipitated with protein A-Sepharose, resolved by SDS-PAGE and visualized by immunoblotting with rabbit anti-FasL antibody (Oncogene Research).

**Analysis of complex formation.** Flag-tagged soluble FasL (25 µg) was incubated with buffer or with DcR3-Fc (40 µg) for 1.5 h at 24 °C. The reaction was loaded onto a Superdex 200 HR 10/30 column (Pharmacia) and developed with PBS; 0.6-ml fractions were collected. The presence of DcR3-Fc-FasL complex in each fraction was analysed by placing 100 µl aliquots into microtitre wells precoated with anti-human IgG (Boehringer) to capture DcR3-Fc, followed by detection with biotinylated anti-Flag antibody Bio M2 (Kodak) and streptavidin-horseradish peroxidase (Amersham). Calibration of the column indicated an apparent relative molecular mass of the complex of 420K (data not shown), which is consistent with a stoichiometry of two DcR3-Fc homodimers to two soluble FasL homotrimers.

**Equilibrium binding analysis.** Microtitre wells were coated with anti-human

IgG, blocked with 2% BSA in PBS. DcR3-Fc or Fas-Fc was added, followed by serially diluted Flag-tagged soluble FasL. Bound ligand was detected with anti-Flag antibody as above. In the competition assay, Fas-Fc was immobilized as above, and the wells were blocked with excess IgG1 before addition of Flag-tagged soluble FasL plus DcR3-Fc.

**T-cell AICD.** CD3<sup>+</sup> lymphocytes were isolated from peripheral blood of individual donors using anti-CD3 magnetic beads (Miltenyi Biotech), stimulated with phytohaemagglutinin (PHA; 2 µg ml<sup>-1</sup>) for 24 h, and cultured in the presence of interleukin-2 (100 U ml<sup>-1</sup>) for 5 days. The cells were plated in wells coated with anti-CD3 antibody (Pharmingen) and analysed for apoptosis 16 h later by FACS analysis of annexin-V-binding of CD4<sup>+</sup> cells<sup>24</sup>.

**Natural killer cell activity.** Natural killer cells were isolated from peripheral blood of individual donors using anti-CD56 magnetic beads (Miltenyi Biotech), and incubated for 16 h with <sup>51</sup>Cr-loaded Jurkat cells at an effector-to-target ratio of 1:1 in the presence of DcR3-Fc, Fas-Fc or human IgG1. Target-cell death was determined by release of <sup>51</sup>Cr in effector-target co-cultures relative to release of <sup>51</sup>Cr by detergent lysis of equal numbers of Jurkat cells.

**Gene-amplification analysis.** Surgical specimens were provided by J. Kern (lung tumours) and P. Quirke (colon tumours). Genomic DNA was extracted (Qiagen) and the concentration was determined using Hoechst dye 33258 intercalation fluorimetry. Amplification was determined by quantitative PCR<sup>18</sup> using a TaqMan instrument (ABI). The method was validated by comparison of PCR and Southern hybridization data for the Myc and HER-2 oncogenes (data not shown). Gene-specific primers and fluorogenic probes were designed on the basis of the sequence of DcR3 or of nearby regions identified on a BAC carrying the human DcR3 gene; alternatively, primers and probes were based on Stanford Human Genome Center marker AFM218xe7 (T160), which is linked to DcR3 (likelihood score = 5.4), SHGC-36268 (T159), the nearest available marker which maps to ~500 kilobases from T160, and five extra markers that span chromosome 20. The DcR3-specific primer sequences were 5'-CTTCTTCGCGCAGCTG-3' and 5'-ATCAGCCCGGACACAG-3' and the fluorogenic probe sequence was 5'-(FAM-ACACGATGCGTGTCTCAAGCAG AAT-TAMARA), where FAM is 5'-fluorescein phosphoramidite. Relative gene-copy numbers were derived using the formula 2<sup>(ΔCT)</sup>, where ΔCT is the difference in amplification cycles required to detect DcR3 in peripheral blood lymphocyte DNA compared to test DNA.

Received 24 September; accepted 6 November 1998.

- Nagata, S. Apoptosis by death factor. *Cell* 88, 355-365 (1997).
- Smith, C. A., Farrah, T. & Goodwin, R. G. The TNF receptor superfamily of cellular and viral proteins: activation, costimulation, and death. *Cell* 76, 959-962 (1994).
- Simonet, W. S. et al. Osteoprotegerin: a novel secreted protein involved in the regulation of bone density. *Cell* 89, 309-319 (1997).
- Suda, T., Takahashi, T., Golstein, P. & Nagata, S. Molecular cloning and expression of Fas ligand, a novel member of the TNF family. *Cell* 75, 1169-1178 (1993).
- Pennica, D. et al. Human tumour necrosis factor: precursor structure, expression and homology to lymphotoxin. *Nature* 312, 724-729 (1984).
- Pitti, R. M. et al. Induction of apoptosis by Apo-2 ligand, a new member of the tumor necrosis factor receptor family. *J. Biol. Chem.* 271, 12687-12690 (1996).
- Wiley, S. R. et al. Identification and characterization of a new member of the TNF family that induces apoptosis. *Immunity* 3, 673-682 (1995).
- Marsters, S. A. et al. Identification of a ligand for the death-domain-containing receptor Apo3. *Curr. Biol.* 8, 525-528 (1998).
- Chicheportiche, Y. et al. TWEAK, a new secreted ligand in the TNF family that weakly induces apoptosis. *J. Biol. Chem.* 272, 32401-32410 (1997).
- Wong, B. R. et al. TRANCE is a novel ligand of the TNFR family that activates c-Jun-N-terminal kinase in T cells. *J. Biol. Chem.* 272, 25190-25194 (1997).
- Anderson, D. M. et al. A homolog of the TNF receptor and its ligand enhance T-cell growth and dendritic-cell function. *Nature* 390, 175-179 (1997).
- Lacey, D. L. et al. Osteoprotegerin ligand is a cytokine that regulates osteoclast differentiation and activation. *Cell* 93, 165-176 (1998).
- Dhein, J., Walczak, H., Baumler, C., Debatin, K. M. & Krammer, P. H. Autocrine T-cell suicide mediated by Apo1/Fas/CD95. *Nature* 373, 438-441 (1995).
- Arase, H., Arase, N. & Saito, T. Fas-mediated cytotoxicity by freshly isolated natural killer cells. *J. Exp. Med.* 181, 1235-1238 (1995).
- Medvedev, A. E. et al. Regulation of Fas and Fas ligand expression in NK cells by cytokines and the involvement of Fas ligand in NK/LAK cell-mediated cytotoxicity. *Cytokine* 9, 394-404 (1997).
- Moretta, A. Mechanisms in cell-mediated cytotoxicity. *Cell* 90, 13-18 (1997).
- Tanaka, M., Itai, T., Adachi, M. & Nagata, S. Downregulation of Fas ligand by shedding. *Nature Med.* 4, 31-36 (1998).
- Gelmini, S. et al. Quantitative PCR-based homogeneous assay with fluorogenic probes to measure c-erbB-2 oncogene amplification. *Clin. Chem.* 43, 752-758 (1997).
- Emery, J. G. et al. Osteoprotegerin is a receptor for the cytotoxic ligand TRAIL. *J. Biol. Chem.* 273, 14363-14367 (1998).
- Wallach, D. Placing death under control. *Nature* 388, 123-125 (1997).
- Colotta, F. et al. Interleukin-1 type II receptor: a decoy target for IL-1 that is regulated by IL-4. *Science* 261, 472-475 (1993).

22. Ashkenazi, A. & Dixit, V. M. Death receptors: signaling and modulation. *Science* 281, 1305–1308 (1998).
23. Ashkenazi, A. & Chamow, S. M. Immunoadhesins as research tools and therapeutic agents. *Curr. Opin. Immunol.* 9, 195–200 (1997).
24. Masters, S. *et al.* Activation of apoptosis by Apo-2 ligand is independent of FADD but blocked by CrmA. *Curr. Biol.* 6, 750–752 (1996).

Acknowledgements. We thank C. Clark, D. Pennica and V. Dixit for comments, and J. Kern and P. Quirke for tumour specimens.

Correspondence and requests for materials should be addressed to A.A. (e-mail: aa@gene.com). The GenBank accession number for the Dcr3 cDNA sequence is AF104419.

## Crystal structure of the ATP-binding subunit of an ABC transporter

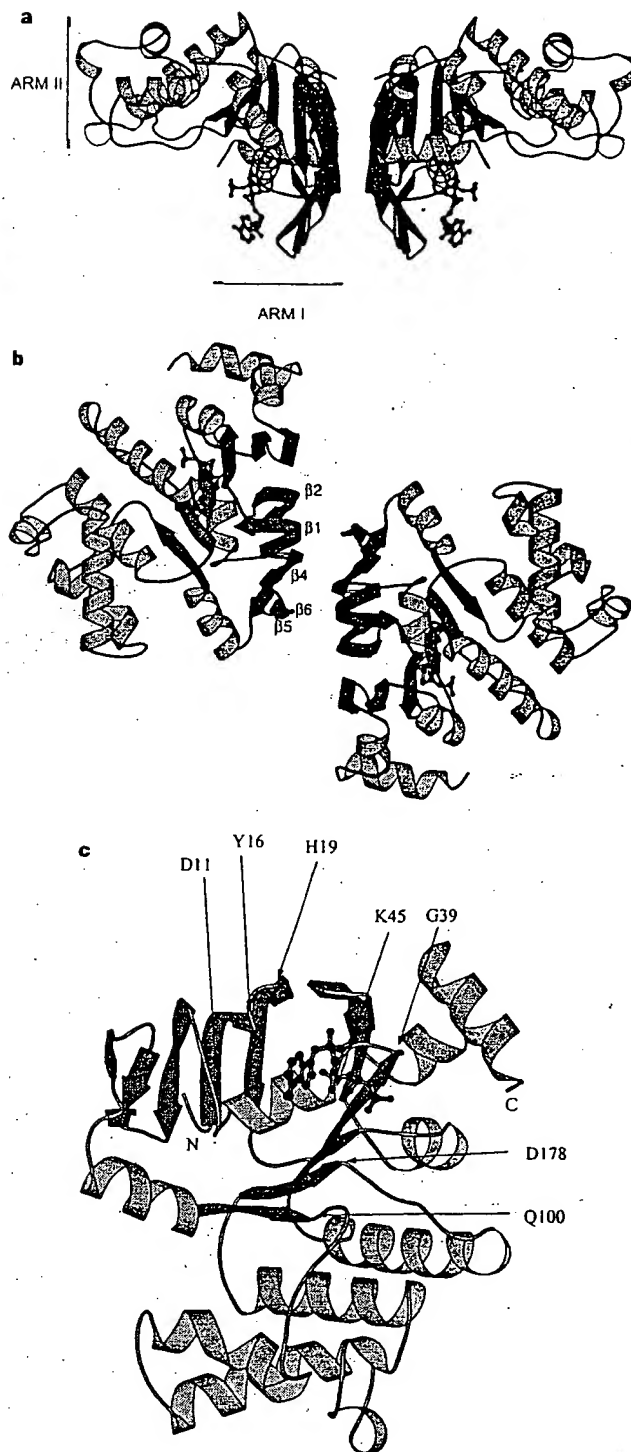
Li-Wei Hung\*, Iris Xiaoyan Wang†, Kishiko Nikaido‡, Pei-Qi Liu†, Giovanna Ferro-Luzzi Ames† & Sung-Hou Kim\*†

\* E. O. Lawrence Berkeley National Laboratory, † Department of Molecular and Cell Biology, and ‡ Department of Chemistry, University of California at Berkeley, Berkeley, California 94720, USA

ABC transporters (also known as traffic ATPases) form a large family of proteins responsible for the translocation of a variety of compounds across membranes of both prokaryotes and eukaryotes<sup>1</sup>. The recently completed *Escherichia coli* genome sequence revealed that the largest family of paralogous *E. coli* proteins is composed of ABC transporters<sup>2</sup>. Many eukaryotic proteins of medical significance belong to this family, such as the cystic fibrosis transmembrane conductance regulator (CFTR), the P-glycoprotein (or multidrug-resistance protein) and the heterodimeric transporter associated with antigen processing (Tap1–Tap2). Here we report the crystal structure at 1.5 Å resolution of HisP, the ATP-binding subunit of the histidine permease, which is an ABC transporter from *Salmonella typhimurium*. We correlate the details of this structure with the biochemical, genetic and biophysical properties of the wild-type and several mutant HisP proteins. The structure provides a basis for understanding properties of ABC transporters and of defective CFTR proteins.

ABC transporters contain four structural domains: two nucleotide-binding domains (NBDs), which are highly conserved throughout the family, and two transmembrane domains<sup>1</sup>. In prokaryotes these domains are often separate subunits which are assembled into a membrane-bound complex; in eukaryotes the domains are generally fused into a single polypeptide chain. The periplasmic histidine permease of *S. typhimurium* and *E. coli*<sup>1,3–5</sup> is a well-characterized ABC transporter that is a good model for this superfamily. It consists of a membrane-bound complex, HisQMP<sub>2</sub>, which comprises integral membrane subunits, HisQ and HisM, and two copies of HisP, the ATP-binding subunit. HisP, which has properties intermediate between those of integral and peripheral membrane proteins<sup>6</sup>, is accessible from both sides of the membrane, presumably by its interaction with HisQ and HisM<sup>6</sup>. The two HisP subunits form a dimer, as shown by their cooperativity in ATP hydrolysis<sup>5</sup>, the requirement for both subunits to be present for activity<sup>7</sup>, and the formation of a HisP dimer upon chemical cross-linking. Soluble HisP also forms a dimer<sup>8</sup>. HisP has been purified and characterized in an active soluble form<sup>3</sup> which can be reconstituted into a fully active membrane-bound complex<sup>9</sup>.

The overall shape of the crystal structure of the HisP monomer is that of an 'L' with two thick arms (arm I and arm II); the ATP-binding pocket is near the end of arm I (Fig. 1). A six-stranded  $\beta$ -sheet ( $\beta$ 3 and  $\beta$ 8– $\beta$ 12) spans both arms of the L, with a domain of  $\alpha$ - plus  $\beta$ -type structure ( $\beta$ 1,  $\beta$ 2,  $\beta$ 4– $\beta$ 7,  $\alpha$ 1 and  $\alpha$ 2) on one side (within arm I) and a domain of mostly  $\alpha$ -helices ( $\alpha$ 3– $\alpha$ 9) on the



**Figure 1** Crystal structure of HisP. **a**, View of the dimer along an axis perpendicular to its two-fold axis. The top and bottom of the dimer are suggested to face towards the periplasmic and cytoplasmic sides, respectively (see text). The thickness of arm II is about 25 Å, comparable to that of membrane.  $\alpha$ -Helices are shown in orange and  $\beta$ -sheets in green. **b**, View along the two-fold axis of the HisP dimer, showing the relative displacement of the monomers not apparent in **a**. The  $\beta$ -strands at the dimer interface are labelled. **c**, View of one monomer from the bottom of arm I, as shown in **a**, towards arm II, showing the ATP-binding pocket. **a–c**, The protein and the bound ATP are in 'ribbon' and 'ball-and-stick' representations, respectively. Key residues discussed in the text are indicated in **c**. These figures were prepared with MOLSCRIPT<sup>28</sup>. N, amino terminus; C, C terminus.

## NOVEL APPROACH TO QUANTITATIVE POLYMERASE CHAIN REACTION USING REAL-TIME DETECTION: APPLICATION TO THE DETECTION OF GENE AMPLIFICATION IN BREAST CANCER

Ivan BIÈCHE<sup>1,2</sup>, Martine OLIVI<sup>1</sup>, Marie-Hélène CHAMPÈME<sup>2</sup>, Dominique VIDAUD<sup>1</sup>, Rosette LIDEREAU<sup>2</sup> and Michel VIDAUD<sup>1\*</sup>

<sup>1</sup>Laboratoire de Génétique Moléculaire, Faculté des Sciences Pharmaceutiques et Biologiques de Paris, Paris, France

<sup>2</sup>Laboratoire d'Oncogénétique, Centre René Huguenin, St-Cloud, France

Gene amplification is a common event in the progression of human cancers, and amplified oncogenes have been shown to have diagnostic, prognostic and therapeutic relevance. A kinetic quantitative polymerase-chain-reaction (PCR) method, based on fluorescent TaqMan methodology and a new instrument (ABI Prism 7700 Sequence Detection System) capable of measuring fluorescence in real-time, was used to quantify gene amplification in tumor DNA. Reactions are characterized by the point during cycling when PCR amplification is still in the exponential phase, rather than the amount of PCR product accumulated after a fixed number of cycles. None of the reaction components is limited during the exponential phase, meaning that values are highly reproducible in reactions starting with the same copy number. This greatly improves the precision of DNA quantification. Moreover, real-time PCR does not require post-PCR sample handling, thereby preventing potential PCR-product carry-over contamination; it possesses a wide dynamic range of quantification and results in much faster and higher sample throughput. The real-time PCR method, was used to develop and validate a simple and rapid assay for the detection and quantification of the 3 most frequently amplified genes (*myc*, *ccnd1* and *erbB2*) in breast tumors. Extra copies of *myc*, *ccnd1* and *erbB2* were observed in 10, 23 and 15%, respectively, of 108 breast-tumor DNA; the largest observed numbers of gene copies were 4.6, 18.6 and 15.1, respectively. These results correlated well with those of Southern blotting. The use of this new semi-automated technique will make molecular analysis of human cancers simpler and more reliable, and should find broad applications in clinical and research settings. *Int. J. Cancer* 78:661–666, 1998.

© 1998 Wiley-Liss, Inc.

Gene amplification plays an important role in the pathogenesis of various solid tumors, including breast cancer, probably because over-expression of the amplified target genes confers a selective advantage. The first technique used to detect genomic amplification was cytogenetic analysis. Amplification of several chromosome regions, visualized either as extrachromosomal double minutes (dmins) or as integrated homogeneously staining regions (HSRs), are among the main visible cytogenetic abnormalities in breast tumors. Other techniques such as comparative genomic hybridization (CGH) (Kallioniemi *et al.*, 1994) have also been used in broad searches for regions of increased DNA copy numbers in tumor cells, and have revealed some 20 amplified chromosome regions in breast tumors. Positional cloning efforts are underway to identify the critical gene(s) in each amplified region. To date, genes known to be amplified frequently in breast cancers include *myc* (8q24), *ccnd1* (11q13), and *erbB2* (17q12-q21) (for review, see Bièche and Lidereau, 1995).

Amplification of the *myc*, *ccnd1*, and *erbB2* proto-oncogenes should have clinical relevance in breast cancer, since independent studies have shown that these alterations can be used to identify sub-populations with a worse prognosis (Berns *et al.*, 1992; Schuuring *et al.*, 1992; Slamon *et al.*, 1987). Muss *et al.* (1994) suggested that these gene alterations may also be useful for the prediction and assessment of the efficacy of adjuvant chemotherapy and hormone therapy.

However, published results diverge both in terms of the frequency of these alterations and their clinical value. For instance, over 500 studies in 10 years have failed to resolve the controversy

surrounding the link suggested by Slamon *et al.* (1987) between *erbB2* amplification and disease progression. These discrepancies are partly due to the clinical, histological and ethnic heterogeneity of breast cancer, but technical considerations are also probably involved.

Specific genes (DNA) were initially quantified in tumor cells by means of blotting procedures such as Southern and slot blotting. These batch techniques require large amounts of DNA (5–10 µg/reaction) to yield reliable quantitative results. Furthermore, meticulous care is required at all stages of the procedures to generate blots of sufficient quality for reliable dosage analysis. Recently, PCR has proven to be a powerful tool for quantitative DNA analysis, especially with minimal starting quantities of tumor samples (small, early-stage tumors and formalin-fixed, paraffin-embedded tissues).

Quantitative PCR can be performed by evaluating the amount of product either after a given number of cycles (end-point quantitative PCR) or after a varying number of cycles during the exponential phase (kinetic quantitative PCR). In the first case, an internal standard distinct from the target molecule is required to ascertain PCR efficiency. The method is relatively easy but implies generating, quantifying and storing an internal standard for each gene studied. Nevertheless, it is the most frequently applied method to date.

One of the major advantages of the kinetic method is its rapidity in quantifying a new gene, since no internal standard is required (an external standard curve is sufficient). Moreover, the kinetic method has a wide dynamic range (at least 5 orders of magnitude), giving an accurate value for samples differing in their copy number. Unfortunately, the method is cumbersome and has therefore been rarely used. It involves aliquot sampling of each assay mix at regular intervals and quantifying, for each aliquot, the amplification product. Interest in the kinetic method has been stimulated by a novel approach using fluorescent TaqMan methodology and a new instrument (ABI Prism 7700 Sequence Detection System) capable of measuring fluorescence in real time (Gibson *et al.*, 1996; Heid *et al.*, 1996). The TaqMan reaction is based on the 5' nuclease assay first described by Holland *et al.* (1991). The latter uses the 5' nuclease activity of Taq polymerase to cleave a specific fluorogenic oligonucleotide probe during the extension phase of PCR. The approach uses dual-labeled fluorogenic hybridization probes (Lee *et al.*, 1993). One fluorescent dye, co-valently linked to the 5' end of the oligonucleotide, serves as a reporter [FAM (*i.e.*, 6-carboxy-fluorescein)] and its emission spectrum is quenched by a second fluorescent dye, TAMRA (*i.e.*, 6-carboxy-tetramethyl-rhodamine) attached to the 3' end. During the extension phase of the PCR

Grant sponsors: Association Pour la Recherche sur le Cancer and Ministère de l'Enseignement Supérieur et de la Recherche.

\*Correspondence to: Laboratoire de Génétique Moléculaire, Faculté des Sciences Pharmaceutiques et Biologiques de Paris, 4 Avenue de l'Observatoire, F-75006 Paris, France. Fax: (33)1-4407-1754. E-mail: mvidauid@teaser.fr

Received 2 May 1998; Revised 30 June 1998

cycle, the fluorescent hybridization probe is hydrolyzed by the 5'-3' nucleolytic activity of DNA polymerase. Nuclease degradation of the probe releases the quenching of FAM fluorescence emission, resulting in an increase in peak fluorescence emission. The fluorescence signal is normalized by dividing the emission intensity of the reporter dye (FAM) by the emission intensity of a reference dye (*i.e.*, ROX, 6-carboxy-X-rhodamine) included in TaqMan buffer, to obtain a ratio defined as the Rn (normalized reporter) for a given reaction tube. The use of a sequence detector enables the fluorescence spectra of all 96 wells of the thermal cycler to be measured continuously during PCR amplification.

The real-time PCR method offers several advantages over other current quantitative PCR methods (Celi *et al.*, 1994): (i) the probe-based homogeneous assay provides a real-time method for detecting only specific amplification products, since specific hybridization of both the primers and the probe is necessary to generate a signal; (ii) the  $C_t$  (threshold cycle) value used for quantification is measured when PCR amplification is still in the log phase of PCR product accumulation. This is the main reason why  $C_t$  is a more reliable measure of the starting copy number than are end-point measurements, in which a slight difference in a limiting component can have a drastic effect on the amount of product; (iii) use of  $C_t$  values gives a wider dynamic range (at least 5 orders of magnitude), reducing the need for serial dilution; (iv) The real-time PCR method is run in a closed-tube system and requires no post-PCR sample handling, thus avoiding potential contamination; (v) the system is highly automated, since the instrument continuously measures fluorescence in all 96 wells of the thermal cycler during PCR amplification and the corresponding software processes, and analyzes the fluorescence data; (vi) the assay is rapid, as results are available just one minute after thermal cycling is complete; (vii) the sample throughput of the method is high, since 96 reactions can be analyzed in 2 hr.

Here, we applied this semi-automated procedure to determine the copy numbers of the 3 most frequently amplified genes in breast tumors (*myc*, *ccnd1* and *erbB2*), as well as 2 genes (*alb* and *app*) located in a chromosome region in which no genetic changes have been observed in breast tumors. The results for 108 breast tumors were compared with previous Southern-blot data for the same samples.

#### MATERIAL AND METHODS

##### Tumor and blood samples

Samples were obtained from 108 primary breast tumors removed surgically from patients at the Centre René Huguénin; none of the patients had undergone radiotherapy or chemotherapy. Immediately after surgery, the tumor samples were placed in liquid nitrogen until extraction of high-molecular-weight DNA. Patients were included in this study if the tumor sample used for DNA preparation contained more than 60% of tumor cells (histological analysis). A blood sample was also taken from 18 of the same patients.

DNA was extracted from tumor tissue and blood leukocytes according to standard methods.

##### Real-time PCR

**Theoretical basis.** Reactions are characterized by the point during cycling when amplification of the PCR product is first detected, rather than by the amount of PCR product accumulated after a fixed number of cycles. The higher the starting copy number of the genomic DNA target, the earlier a significant increase in fluorescence is observed. The parameter  $C_t$  (threshold cycle) is defined as the fractional cycle number at which the fluorescence generated by cleavage of the probe passes a fixed threshold above baseline. The target gene copy number in unknown samples is quantified by measuring  $C_t$  and by using a standard curve to determine the starting copy number. The precise amount of genomic DNA (based on optical density) and its quality (*i.e.*, lack

of extensive degradation) are both difficult to assess. We therefore also quantified a control gene (*alb*) mapping to chromosome region 4q11-q13, in which no genetic alterations have been found in breast-tumor DNA by means of CGH (Kallioniemi *et al.*, 1994).

Thus, the ratio of the copy number of the target gene to the copy number of the *alb* gene normalizes the amount and quality of genomic DNA. The ratio defining the level of amplification is termed "N", and is determined as follows:

$$N = \frac{\text{copy number of target gene (app, myc, ccnd1, erbB2)}}{\text{copy number of reference gene (alb)}}$$

**Primers, probes, reference human genomic DNA and PCR consumables.** Primers and probes were chosen with the assistance of the computer programs Oligo 4.0 (National Biosciences, Plymouth, MN), EuGene (Daniben Systems, Cincinnati, OH) and Primer Express (Perkin-Elmer Applied Biosystems, Foster City, CA).

Primers were purchased from DNAgency (Malvern, PA) and probes from Perkin-Elmer Applied Biosystems.

Nucleotide sequences for the oligonucleotide hybridization probes and primers are available on request.

The TaqMan PCR Core reagent kit, MicroAmp optical tubes, and MicroAmp caps were from Perkin-Elmer Applied Biosystems.

**Standard-curve construction.** The kinetic method requires a standard curve. The latter was constructed with serial dilutions of specific PCR products, according to Piatk *et al.* (1993). In practice, each specific PCR product was obtained by amplifying 20 ng of a standard human genomic DNA (Boehringer, Mannheim, Germany) with the same primer pairs as those used later for real-time quantitative PCR. The 5 PCR products were purified using MicroSpin S-400 HR columns (Pharmacia, Uppsala, Sweden) electrophoresed through an acrylamide gel and stained with ethidium bromide to check their quality. The PCR products were then quantified spectrophotometrically and pooled, and serially diluted 10-fold in mouse genomic DNA (Clontech, Palo Alto, CA) at a constant concentration of 2 ng/ $\mu$ l. The standard curve used for real-time quantitative PCR was based on serial dilutions of the pool of PCR products ranging from  $10^{-7}$  ( $10^5$  copies of each gene) to  $10^{-10}$  ( $10^2$  copies). This series of diluted PCR products was aliquoted and stored at  $-80^\circ\text{C}$  until use.

The standard curve was validated by analyzing 2 known quantities of calibrator human genomic DNA (20 ng and 50 ng).

**PCR amplification.** Amplification mixes (50  $\mu$ l) contained the sample DNA (around 20 ng, around 6600 copies of disomic genes),  $10\times$  TaqMan buffer (5  $\mu$ l), 200  $\mu$ M dATP, dCTP, dGTP, and 400  $\mu$ M dUTP, 5 mM  $\text{MgCl}_2$ , 1.25 units of AmpliTaq Gold, 0.5 units of AmpErase uracil N-glycosylase (UNG), 200 nM each primer and 100 nM probe. The thermal cycling conditions comprised 2 min at  $50^\circ\text{C}$  and 10 min at  $95^\circ\text{C}$ . Thermal cycling consisted of 40 cycles at  $95^\circ\text{C}$  for 15 s and  $65^\circ\text{C}$  for 1 min. Each assay included: a standard curve (from  $10^5$  to  $10^2$  copies) in duplicate, a no-template control, 20 ng and 50 ng of calibrator human genomic DNA (Boehringer) in triplicate, and about 20 ng of unknown genomic DNA in triplicate (26 samples can thus be analyzed on a 96-well microplate). All samples with a coefficient of variation (CV) higher than 10% were retested.

All reactions were performed in the ABI Prism 7700 Sequence Detection System (Perkin-Elmer Applied Biosystems), which detects the signal from the fluorogenic probe during PCR.

**Equipment for real-time detection.** The 7700 system has a built-in thermal cycler and a laser directed via fiber optical cables to each of the 96 sample wells. A charge-coupled-device (CDD) camera collects the emission from each sample and the data are analyzed automatically. The software accompanying the 7700 system calculates  $C_t$  and determines the starting copy number in the samples.

**Determination of gene amplification.** Gene amplification was calculated as described above. Only samples with an N value higher than 2 were considered to be amplified.

# RESULTS

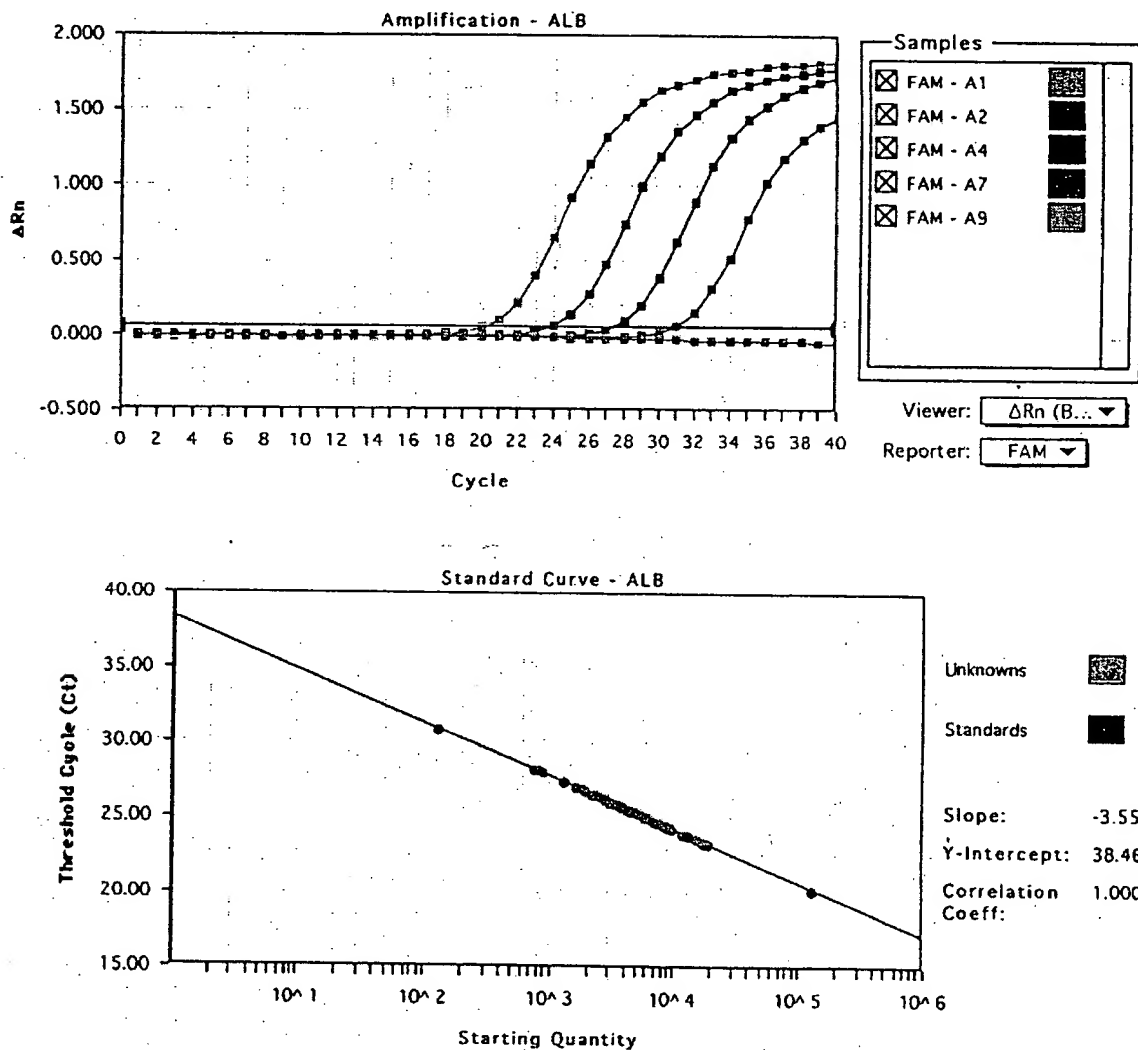
To validate the method, real-time PCR was performed on genomic DNA extracted from 108 primary breast tumors, and 18 normal leukocyte DNA samples from some of the same patients. The target genes were the *myc*, *ccnd1* and *erbB2* proto-oncogenes, and the  $\beta$ -amyloid precursor protein gene (*app*), which maps to a chromosome region (21q21.2) in which no genetic alterations have been found in breast tumors (Kallioniemi *et al.*, 1994). The reference disomic gene was the albumin gene (*alb*, chromosome 4q11-q13).

## Validation of the standard curve and dynamic range of real-time PCR

The standard curve was constructed from PCR products serially diluted in genomic mouse DNA at a constant concentration of 2 ng/ $\mu$ l. It should be noted that the 5 primer pairs chosen to analyze the 5 target genes do not amplify genomic mouse DNA (data not shown). Figure 1 shows the real-time PCR standard curve for the *alb* gene. The dynamic range was wide (at least 4 orders of magnitude), with samples containing as few as  $10^2$  copies or as many as  $10^5$  copies.

## Copy-number ratio of the 2 reference genes (*app* and *alb*)

The *app* to *alb* copy-number ratio was determined in 18 normal leukocyte DNA samples and all 108 primary breast-tumor DNA



**FIGURE 1** – Albumin (*alb*) gene dosage by real-time PCR. Top: Amplification plots for reactions with starting *alb* gene copy number ranging from  $10^5$  (A9),  $10^4$  (A7),  $10^3$  (A4) to  $10^2$  (A2) and a no-template control (A1). Cycle number is plotted vs. change in normalized reporter signal ( $\Delta Rn$ ). For each reaction tube, the fluorescence signal of the reporter dye (FAM) is divided by the fluorescence signal of the passive reference dye (ROX), to obtain a ratio defined as the normalized reporter signal (Rn).  $\Delta Rn$  represents the normalized reporter signal (Rn) minus the baseline signal established in the first 15 PCR cycles.  $\Delta Rn$  increases during PCR as *alb* PCR product copy number increases until the reaction reaches a plateau.  $C_t$  (threshold cycle) represents the fractional cycle number at which a significant increase in Rn above a baseline signal (horizontal black line) can first be detected. Two replicate plots were performed for each standard sample, but the data for only one are shown here. Bottom: Standard curve plotting log starting copy number vs.  $C_t$  (threshold cycle). The black dots represent the data for standard samples plotted in duplicate and the red dots the data for unknown genomic DNA samples plotted in triplicate. The standard curve shows 4 orders of linear dynamic range.

samples. We selected these 2 genes because they are located in 2 chromosome regions (*app*, 21q21.2; *alb*, 4q11-q13) in which no obvious genetic changes (including gains or losses) have been observed in breast cancers (Kallioniemi *et al.*, 1994). The ratio for the 18 normal leukocyte DNA samples fell between 0.7 and 1.3 (mean  $1.02 \pm 0.21$ ), and was similar for the 108 primary breast-tumor DNA samples (0.6 to 1.6, mean  $1.06 \pm 0.25$ ), confirming that *alb* and *app* are appropriate reference disomic genes for breast-tumor DNA. The low range of the ratios also confirmed that the nucleotide sequences chosen for the primers and probes were not polymorphic, as mismatches of their primers or probes with the subject's DNA would have resulted in differential amplification.

#### *myc*, *ccnd1* and *erbB2* gene dose in normal leukocyte DNA

To determine the cut-off point for gene amplification in breast-cancer tissue, 18 normal leukocyte DNA samples were tested for the gene dose (N), calculated as described in "Material and Methods". The N value of these samples ranged from 0.5 to 1.3 (mean  $0.84 \pm 0.22$ ) for *myc*; 0.7 to 1.6 (mean  $1.06 \pm 0.23$ ) for *ccnd1* and 0.6 to 1.3 (mean  $0.91 \pm 0.19$ ) for *erbB2*. Since N values for *myc*, *ccnd1* and *erbB2* in normal leukocyte DNA consistently fell between 0.5 and 1.6, values of 2 or more were considered to represent gene amplification in tumor DNA.

#### *myc*, *ccnd1* and *erbB2* gene dose in breast-tumor DNA

*myc*, *ccnd1* and *erbB2* gene copy numbers in the 108 primary breast tumors are reported in Table I. Extra copies of *ccnd1* were more frequent (23%, 25/108) than extra copies of *erbB2* (15%, 16/108) and *myc* (10%, 11/108), and ranged from 2 to 18.6 for *ccnd1*, 2 to 15.1 for *erbB2*, and only 2 to 4.6 for the *myc* gene. Figure 2 and Table II represent tumors in which the *ccnd1* gene was amplified 16-fold (T145), 6-fold (T133) and non-amplified (T118). The 3 genes were never found to be co-amplified in the same tumor. *erbB2* and *ccnd1* were co-amplified in only 3 cases, *myc* and *ccnd1* in 2 cases and *myc* and *erbB2* in 1 case. This favors the hypothesis that gene amplifications are independent events in breast cancer. Interestingly, 5 tumors showed a decrease of at least 50% in the *erbB2* copy number ( $N < 0.5$ ), suggesting that they bore deletions of the 17q21 region (the site of *erbB2*). No such decrease in copy number was observed with the other 2 proto-oncogenes.

#### Comparison of gene dose determined by real-time quantitative PCR and Southern-blot analysis

Southern-blot analysis of *myc*, *ccnd1* and *erbB2* amplifications had previously been done on the same 108 primary breast tumors. A perfect correlation between the results of real-time PCR and Southern blot was obtained for tumors with high copy numbers ( $N \geq 5$ ). However, there were cases (1 *myc*, 6 *ccnd1* and 4 *erbB2*) in which real-time PCR showed gene amplification whereas Southern-blot did not, but these were mainly cases with low extra copy numbers (N from 2 to 2.9).

### DISCUSSION

The clinical applications of gene amplification assays are currently limited, but would certainly increase if a simple, standardized and rapid method were perfected. Gene amplification status has been studied mainly by means of Southern blotting, but this method is not sensitive enough to detect low-level gene amplification nor accurate enough to quantify the full range of amplification values. Southern blotting is also time-consuming, uses radioactive

reagents and requires relatively large amounts of high-quality genomic DNA, which means it cannot be used routinely in many laboratories. An amplification step is therefore required to determine the copy number of a given target gene from minimal quantities of tumor DNA (small early-stage tumors, cytopuncture specimens or formalin-fixed, paraffin-embedded tissues).

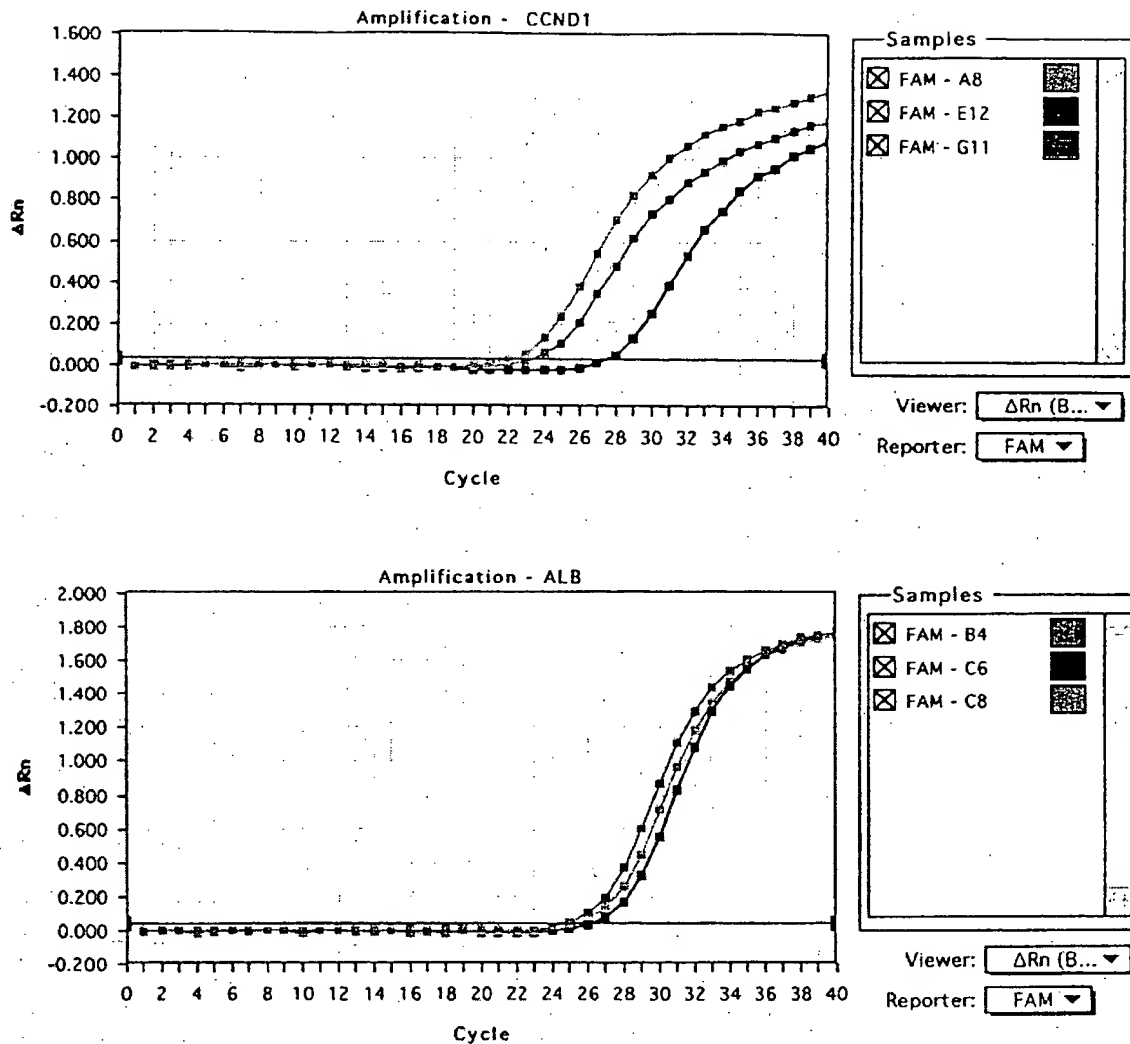
In this study, we validated a PCR method developed for the quantification of gene over-representation in tumors. The method, based on real-time analysis of PCR amplification, has several advantages over other PCR-based quantitative assays such as competitive quantitative PCR (Celi *et al.*, 1994). First, the real-time PCR method is performed in a closed-tube system, avoiding the risk of contamination by amplified products. Re-amplification of carryover PCR products in subsequent experiments can also be prevented by using the enzyme uracil N-glycosylase (UNG) (Longo *et al.*, 1990). The second advantage is the simplicity and rapidity of sample analysis, since no post-PCR manipulations are required. Our results show that the automated method is reliable. We found it possible to determine, in triplicate, the number of copies of a target gene in more than 100 tumors per day. Third, the system has a linear dynamic range of at least 4 orders of magnitude, meaning that samples do not have to contain equal starting amounts of DNA. This technique should therefore be suitable for analyzing formalin-fixed, paraffin-embedded tissues. Fourth, and above all, real-time PCR makes DNA quantification much more precise and reproducible, since it is based on  $C_t$  values rather than end-point measurement of the amount of accumulated PCR product. Indeed, the ABI-Prism 7700 Sequence Detection System enables  $C_t$  to be calculated when PCR amplification is still in the exponential phase and when none of the reaction components is rate-limiting. The within-run CV of the  $C_t$  value for calibrator human DNA (5 replicates) was always below 5%, and the between-assay precision in 5 different runs was always below 10% (data not shown). In addition, the use of a standard curve is not absolutely necessary, since the copy number can be determined simply by comparing the  $C_t$  ratio of the target gene with that of reference genes. The results obtained by the 2 methods (with and without a standard curve) are similar in our experiments (data not shown). Moreover, unlike competitive quantitative PCR, real-time PCR does not require an internal control (the design and storage of internal controls and the validation of their amplification efficiency is laborious).

The only potential disadvantage of real-time PCR, like all other PCR-based methods and solid-matrix blotting techniques (Southern blots and dot blots) is that it cannot avoid dilution artifacts inherent in the extraction of DNA from tumor cells contained in heterogeneous tissue specimens. Only FISH and immunohistochemistry can measure alterations on a cell-by-cell basis (Pauletti *et al.*, 1996; Slamon *et al.*, 1989). However, FISH requires expensive equipment and trained personnel and is also time-consuming. Moreover, FISH does not assess gene expression and therefore cannot detect cases in which the gene product is over-expressed in the absence of gene amplification, which will be possible in the future by real-time quantitative RT-PCR. Immunohistochemistry is subject to considerable variations in the hands of different teams, owing to alterations of target proteins during the procedure, the different primary antibodies and fixation methods used and the criteria used to define positive staining.

The results of this study are in agreement with those reported in the literature. (i) Chromosome regions 4q11-q13 and 21q21.2 (which bear *alb* and *app*, respectively) showed no genetic alterations in the breast-cancer samples studied here, in keeping with the results of CGH (Kallioniemi *et al.*, 1994). (ii) We found that amplifications of these 3 oncogenes were independent events, as reported by other teams (Berns *et al.*, 1992; Borg *et al.*, 1992). (iii) The frequency and degree of *myc* amplification in our breast tumor DNA series were lower than those of *ccnd1* and *erbB2* amplification, confirming the findings of Borg *et al.* (1992) and Courjal *et al.* (1997). (iv) The maxima of *ccnd1* and *erbB2* over-representation were 18-fold and 15-fold, also in keeping with earlier results (about

TABLE I - DISTRIBUTION OF AMPLIFICATION LEVEL (N) FOR *myc*, *ccnd1* AND *erbB2* GENES IN 108 HUMAN BREAST TUMORS

Gene	Amplification level (N)			
	<0.5	0.5-1.9	2-4.9	$\geq 5$
<i>myc</i>	0	97 (89.8%)	11 (10.2%)	0
<i>ccnd1</i>	0	83 (76.9%)	17 (15.7%)	8 (7.4%)
<i>erbB2</i>	5 (4.6%)	87 (80.6%)	8 (7.4%)	8 (7.4%)



Tumor	CCND1		ALB	
	$C_t$	Copy number	$C_t$	Copy number
T118	27.3	4605	26.5	4365
T133	23.2	61659	25.2	10092
T145	22.1	125892	25.6	7762

FIGURE 2 - *ccnd1* and *alb* gene dosage by real-time PCR in 3 breast tumor samples: T118 (E12, C6, black squares), T133 (G11, B4, red squares) and T145 (A8, C8, blue squares). Given the  $C_t$  of each sample, the initial copy number is inferred from the standard curve obtained during the same experiment. Triplicate plots were performed for each tumor sample, but the data for only one are shown here. The results are shown in Table II.

30-fold maximum) (Berns *et al.*, 1992; Borg *et al.*, 1992; Courjal *et al.*, 1997). (v) The *erbB2* copy numbers obtained with real-time PCR were in good agreement with data obtained with other quantitative PCR-based assays in terms of the frequency and degree of amplification (An *et al.*, 1995; Deng *et al.*, 1996; Valeron

*et al.*, 1996). Our results also correlate well with those recently published by Gelmini *et al.* (1997), who used the TaqMan system to measure *erbB2* amplification in a small series of breast tumors ( $n = 25$ ), but with an instrument (LS-50B luminescence spectrometer, Perkin-Elmer Applied Biosystems) which only allows end-



TABLE II - EXAMPLES OF *ccnd1* GENE DOSAGE RESULTS FROM 3 BREAST TUMORS<sup>1</sup>

Tumor	<i>ccnd1</i>			<i>alb</i>			<i>Nccnd1/alb</i>
	Copy number	Mean	SD	Copy number	Mean	SD	
T118	4525	4603	77	4223	4325	89	1.06
	4605			4365			
	4678			4387			
T133	59821	61100	1111	9787	10137	375	6.03
	61659			10092			
	61821			10533			
T145	128563	125392	3448	7321	7672	316	16.34
	125892			7762			
	121722			7933			

<sup>1</sup>For each sample, 3 replicate experiments were performed and the mean and the standard deviation (SD) was determined. The level of *ccnd1* gene amplification (*Nccnd1/alb*) is determined by dividing the average *ccnd1* copy number value by the average *alb* copy number value.

point measurement of fluorescence intensity. Here we report *myc* and *ccnd1* gene dosage in breast cancer by means of quantitative PCR. (vi) We found a high degree of concordance between real-time quantitative PCR and Southern blot analysis in terms of gene amplification, especially for samples with high copy numbers ( $\geq 5$ -fold). The slightly higher frequency of gene amplification (especially *ccnd1* and *erbB2*) observed by means of real-time quantitative PCR as compared with Southern-blot analysis may be explained by the higher sensitivity of the former method. However, we cannot rule out the possibility that some tumors with a few extra

gene copies observed in real-time PCR had additional copies of an arm or a whole chromosome (trisomy, tetrasomy or polysomy) rather than true gene amplification. These 2 types of genetic alteration (polysomy and gene amplification) could be easily distinguished in the future by using an additional probe located on the same chromosome arm, but some distance from the target gene. It is noteworthy that high gene copy numbers have the greatest prognostic significance in breast carcinoma (Borg *et al.*, 1992; Slamon *et al.*, 1987).

Finally, this technique can be applied to the detection of gene deletion as well as gene amplification. Indeed, we found a decreased copy number of *erbB2* (but not of the other 2 proto-oncogenes) in several tumors; *erbB2* is located in a chromosome region (17q21) reported to contain both deletions and amplifications in breast cancer (Bièche and Lidereau, 1995).

In conclusion, gene amplification in various cancers can be used as a marker of pre-neoplasia, also for early diagnosis of cancer, staging, prognostication and choice of treatment. Southern blotting is not sufficiently sensitive, and FISH is lengthy and complex. Real-time quantitative PCR overcomes both these limitations, and is a sensitive and accurate method of analyzing large numbers of samples in a short time. It should find a place in routine clinical gene dosage.

#### ACKNOWLEDGEMENTS

RL is a research director at the Institut National de la Santé et de la Recherche Médicale (INSERM). We thank the staff of the Centre René Huguenin for assistance in specimen collection and patient care.

#### REFERENCES

- AN, H.X., NIEDERACHER, D., BECKMANN, M.W., GÖHRING, U.J., SCHARL, A., PICARD, F., VAN ROEVEN, C., SCHNÜRCH, H.G. and BENDER, H.G., *erbB2* gene amplification detected by fluorescent differential polymerase chain reaction in paraffin-embedded breast carcinoma tissues. *Int. J. Cancer (Pred. Oncol.)*, 64, 291-297 (1995).
- BERNS, E.M.J.J., KLIJN, J.G.M., VAN PUTTEN, W.L.J., VAN STAVEREN, I.L., PORTINGEN, H. and FOEKENS, J.A., *c-myc* amplification is a better prognostic factor than *HER2/neu* amplification in primary breast cancer. *Cancer Res.*, 52, 1107-1113 (1992).
- BIÈCHE, I. and LIDEREAU, R., Genetic alterations in breast cancer. *Genes Chrom. Cancer*, 14, 227-251 (1995).
- BORG, A., BALDETORP, B., FERNO, M., OLSSON, H. and SIGURDSSON, H., *c-myc* amplification is an independent prognostic factor in post-menopausal breast cancer. *Int. J. Cancer*, 51, 687-691 (1992).
- CÉLI, F.S., COHEN, M.M., ANTONARAKIS, S.E., WERTHEIMER, E., ROTH, J. and SHULDNER, A.R., Determination of gene dosage by a quantitative adaptation of the polymerase chain reaction (q-PCR): rapid detection of deletions and duplications of gene sequences. *Genomics*, 21, 304-310 (1994).
- COURJAL, F., CUNY, M., SIMONY-LAFONTAINE, J., LOUASSON, G., SPEISER, P., ZEILLINGER, R., RODRIGUEZ, C. and THEILLET, C., Mapping of DNA amplifications at 15 chromosomal localizations in 1875 breast tumors: definition of phenotypic groups. *Cancer Res.*, 57, 4360-4367 (1997).
- DENG, G., YU, M., CHEN, L.C., MOORE, D., KURISU, W., KALLIONIEMI, A., WALDMAN, F.M., COLLINS, C. and SMITH, H.S., Amplifications of oncogene *erbB-2* and chromosome 20q in breast cancer determined by differentially competitive polymerase chain reaction. *Breast Cancer Res. Treat.*, 40, 271-281 (1996).
- GELMINI, S., ORIANDO, C., SESTINI, R., VONA, G., PINZANI, P., RUOCCO, L. and PAZZAGLI, M., Quantitative polymerase chain reaction-based homogeneous assay with fluorogenic probes to measure *c-erbB-2* oncogene amplification. *Clin. Chem.*, 43, 752-758 (1997).
- GIBSON, U.E.M., HEID, C.A. and WILLIAMS, P.M., A novel method for real-time quantitative RT-PCR. *Genome Res.*, 6, 995-1001 (1996).
- HEID, C.A., STEVENS, J., LIVAK, K.J. and WILLIAMS, P.M., Real-time quantitative PCR. *Genome Res.*, 6, 986-994 (1996).
- HOLLAND, P.M., ABRAMSON, R.D., WATSON, R. and GELFAND, D.H., Detection of specific polymerase chain reaction product by utilizing the 5' to 3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. nat. Acad. Sci. (Wash.)*, 88, 7276-7280 (1991).
- KALLIONIEMI, A., KALLIONIEMI, O.P., PIPER, J., TANNER, M., STOKKES, T., CHEN, L., SMITH, H.S., PINKEL, D., GRAY, J.W. and WALDMAN, F.M., Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. *Proc. nat. Acad. Sci. (Wash.)*, 91, 2156-2160 (1994).
- LEE, L.G., CONNELL, C.R. and BIOCH, W., Allelic discrimination by nick-translation PCR with fluorogenic probe. *Nucleic Acids Res.*, 21, 3761-3766 (1993).
- LONGO, N., BERNINGER, N.S. and HARTLEY, J.L., Use of uracil DNA glycosylase to control carry-over contamination in polymerase chain reactions. *Gene*, 93, 125-128 (1990).
- MUSS, H.B., THOR, A.D., BERRY, D.A., KUTE, T., LIU, E.T., KOERNER, F., CIRINCIONE, C.T., BUDMAN, D.R., WOOD, W.C., BARCOS, M. and HENDERSON, I.C., *c-erbB-2* expression and response to adjuvant therapy in women with node-positive early breast cancer. *New Engl. J. Med.*, 330, 1260-1266 (1994).
- PAULETTI, G., GODOLPHIN, W., PRESS, M.F. and SALMON, D.J., Detection and quantification of *HER-2/neu* gene amplification in human breast cancer archival material using fluorescence *in situ* hybridization. *Oncogene*, 13, 63-72 (1996).
- PIATAK, M., LUK, K.C., WILLIAMS, B. and LIFSON, J.D., Quantitative competitive polymerase chain reaction for accurate quantitation of HIV DNA and RNA species. *Biotechniques*, 14, 70-80 (1993).
- SCHUURING, E., VERHOEVEN, E., VAN TINTEREN, H., PETERSE, J.L., NUNNIK, B., THUNNISSEN, F.B.J.M., DEVILLE, P., CORNELISSE, C.J., VAN DE VIVER, M.J., MOOI, W.J. and MICHALIDES, R.J.A.M., Amplification of genes within the chromosome 11q13 region is indicative of poor prognosis in patients with operable breast cancer. *Cancer Res.*, 52, 5229-5234 (1992).
- SLAMON, D.J., CLARK, G.M., WONG, S.G., LEVIN, W.S., ULLRICH, A. and MCGUIRE, W.L., Human breast cancer: correlation of relapse and survival with amplification of the *HER-2/neu* oncogene. *Science*, 235, 177-182 (1987).
- SLAMON, D.J., GODOLPHIN, W., JONES, L.A., HOLT, J.A., WONG, S.G., KEITH, D.E., LEVIN, W.J., STUART, S.G., UDVOE, J., ULLRICH, A. and PRESS, M.F., Studies of the *HER-2/neu* proto-oncogene in human breast and ovarian cancer. *Science*, 244, 707-712 (1989).
- VALERON, P.F., CHIRINO, R., FERNANDEZ, L., TORRES, S., NAVARRO, D., AGUIAR, J., CABRERA, J.J., DIAZ-CHICO, B.N. and DIAZ-CHICO, J.C., Validation of a differential PCR and an ELISA procedure in studying *HER-2/neu* status in breast cancer. *Int. J. Cancer*, 65, 129-133 (1996).



IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant : Ashkenazi et al.  
App. No. : 09/903,925  
Filed : July 11, 2001  
For : SECRETED AND  
TRANSMEMBRANE  
POLYPEPTIDES AND NUCLEIC  
ACIDS ENCODING THE SAME  
Examiner : Hamud, Fozia M

Group Art Unit 1647

CERTIFICATE OF EXPRESS MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to Commissioner of Patents, Washington D.C. 20231 on:

(Date)

Commissioner of Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

DECLARATION OF AVI ASHKENAZI, Ph.D UNDER 37 C.F.R. § 1.132

I, Avi Ashkenazi, Ph.D. declare and say as follows: -

1. I am Director and Staff Scientist at the Molecular Oncology Department of Genentech, Inc., South San Francisco, CA 94080.
2. I joined Genentech in 1988 as a postdoctoral fellow. Since then, I have investigated a variety of cellular signal transduction mechanisms, including apoptosis, and have developed technologies to modulate such mechanisms as a means of therapeutic intervention in cancer and autoimmune disease. I am currently involved in the investigation of a series of secreted proteins over-expressed in tumors, with the aim to identify useful targets for the development of therapeutic antibodies for cancer treatment.
3. My scientific Curriculum Vitae, including my list of publications, is attached to and forms part of this Declaration (Exhibit A).
4. Gene amplification is a process in which chromosomes undergo changes to contain multiple copies of certain genes that normally exist as a single copy, and is an important factor in the pathophysiology of cancer. Amplification of certain genes (e.g., Myc or Her2/Neu)

gives cancer cells a growth or survival advantage relative to normal cells, and might also provide a mechanism of tumor cell resistance to chemotherapy or radiotherapy.

5. If gene amplification results in over-expression of the mRNA and the corresponding gene product, then it identifies that gene product as a promising target for cancer therapy, for example by the therapeutic antibody approach. Even in the absence of over-expression of the gene product, amplification of a cancer marker gene - as detected, for example, by the reverse transcriptase TaqMan<sup>®</sup> PCR or the fluorescence *in situ* hybridization (FISH) assays - is useful in the diagnosis or classification of cancer, or in predicting or monitoring the efficacy of cancer therapy. An increase in gene copy number can result not only from intrachromosomal changes but also from chromosomal aneuploidy. It is important to understand that detection of gene amplification can be used for cancer diagnosis even if the determination includes measurement of chromosomal aneuploidy. Indeed, as long as a significant difference relative to normal tissue is detected, it is irrelevant if the signal originates from an increase in the number of gene copies per chromosome and/or an abnormal number of chromosomes.

6. I understand that according to the Patent Office, absent data demonstrating that the increased copy number of a gene in certain types of cancer leads to increased expression of its product, gene amplification data are insufficient to provide substantial utility or well established utility for the gene product (the encoded polypeptide), or an antibody specifically binding the encoded polypeptide. However, even when amplification of a cancer marker gene does not result in significant over-expression of the corresponding gene product, this very absence of gene product over-expression still provides significant information for cancer diagnosis and treatment. Thus, if over-expression of the gene product does not parallel gene amplification in certain tumor types but does so in others, then parallel monitoring of gene amplification and gene product over-expression enables more accurate tumor classification and hence better determination of suitable therapy. In addition, absence of over-expression is crucial information for the practicing clinician. If a gene is amplified but the corresponding gene product is not over-expressed, the clinician accordingly will decide not to treat a patient with agents that target that gene product.

7. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so

made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

By: Avi Ashkenazi  
Avi Ashkenazi, Ph.D.

Date: 9/15/03

SV 455281 v1  
9/12/03 3:06 PM (39780.7000)

## **CURRICULUM VITAE**

**Avi Ashkenazi**

July 2003

### **Personal:**

Date of birth: 29 November, 1956  
Address: 1456 Tarrytown Street, San Mateo, CA 94402  
Phone: (650) 578-9199 (home); (650) 225-1853 (office)  
Fax: (650) 225-6443 (office)  
Email: aa@gene.com

### **Education:**

1983: B.S. in Biochemistry, with honors, Hebrew University, Israel  
1986: Ph.D. in Biochemistry, Hebrew University, Israel

### **Employment:**

1983-1986: Teaching assistant, undergraduate level course in Biochemistry  
1985-1986: Teaching assistant, graduate level course on Signal Transduction  
1986 - 1988: Postdoctoral fellow, Hormone Research Dept., UCSF, and  
Developmental Biology Dept., Genentech, Inc., with J. Ramachandran  
1988 - 1989: Postdoctoral fellow, Molecular Biology Dept., Genentech, Inc.,  
with D. Capon  
1989 - 1993: Scientist, Molecular Biology Dept., Genentech, Inc.  
1994 -1996: Senior Scientist, Molecular Oncology Dept., Genentech, Inc.  
1996-1997: Senior Scientist and Interim director, Molecular Oncology Dept.,  
Genentech, Inc.  
1997-1990: Senior Scientist and preclinical project team leader, Genentech, Inc.  
1999 -2002: Staff Scientist in Molecular Oncology, Genentech, Inc.  
2002-present: Staff Scientist and Director in Molecular Oncology, Genentech, Inc.

### **Awards:**

1988: First prize, The Boehringer Ingelheim Award

## Editorial:

Editorial Board Member: Current Biology

Associate Editor, Clinical Cancer Research.

Associate Editor, Cancer Biology and Therapy.

## Refereed papers:

1. Gertler, A., Ashkenazi, A., and Madar, Z. Binding sites for human growth hormone and ovine and bovine prolactins in the mammary gland and liver of the lactating cow. *Mol. Cell. Endocrinol.* 34, 51-57 (1984).
2. Gertler, A., Shamay, A., Cohen, N., Ashkenazi, A., Friesen, H., Levanon, A., Gorecki, M., Aviv, H., Hadari, D., and Vogel, T. Inhibition of lactogenic activities of ovine prolactin and human growth hormone (hGH) by a novel form of a modified recombinant hGH. *Endocrinology* 118, 720-726 (1986).
3. Ashkenazi, A., Madar, Z., and Gertler, A. Partial purification and characterization of bovine mammary gland prolactin receptor. *Mol. Cell. Endocrinol.* 50, 79-87 (1987).
4. Ashkenazi, A., Pines, M., and Gertler, A. Down-regulation of lactogenic hormone receptors in Nb2 lymphoma cells by cholera toxin. *Biochemistry Internatl.* 14, 1065-1072 (1987).
5. Ashkenazi, A., Cohen, R., and Gertler, A. Characterization of lactogen receptors in lactogenic hormone-dependent and independent Nb2 lymphoma cell lines. *FEBS Lett.* 210, 51-55 (1987).
6. Ashkenazi, A., Vogel, T., Barash, I., Hadari, D., Levanon, A., Gorecki, M., and Gertler, A. Comparative study on in vitro and in vivo modulation of lactogenic and somatotrophic receptors by native human growth hormone and its modified recombinant analog. *Endocrinology* 121, 414-419 (1987).
7. Peralta, E., Winslow, J., Peterson, G., Smith, D., Ashkenazi, A., Ramachandran, J., Schimerlik, M., and Capon, D. Primary structure and biochemical properties of an M2 muscarinic receptor. *Science* 236, 600-605 (1987).
8. Peralta, E., Ashkenazi, A., Winslow, J., Smith, D., Ramachandran, J., and Capon, D. J. Distinct primary structures, ligand-binding properties and tissue-specific expression of four human muscarinic acetylcholine receptors. *EMBO J.* 6, 3923-3929 (1987).
9. Ashkenazi, A., Winslow, J., Peralta, E., Peterson, G., Schimerlik, M., Capon, D., and Ramachandran, J. An M2 muscarinic receptor subtype coupled to both adenylyl cyclase and phosphoinositide turnover. *Science* 238, 672-675 (1987).

10. Pines, M., Ashkenazi, A., Cohen-Chapnik, N., Binder, L., and Gertler, A. Inhibition of the proliferation of Nb2 lymphoma cells by femtomolar concentrations of cholera toxin and partial reversal of the effect by 12-o-tetradecanoyl-phorbol-13-acetate. *J. Cell. Biochem.* 37, 119-129 (1988).
11. Peralta, E. Ashkenazi, A., Winslow, J., Ramachandran, J., and Capon, D. Differential regulation of PI hydrolysis and adenylyl cyclase by muscarinic receptor subtypes. *Nature* 334, 434-437 (1988).
12. Ashkenazi, A., Peralta, E., Winslow, J., Ramachandran, J., and Capon, D. Functionally distinct G proteins couple different receptors to PI hydrolysis in the same cell. *Cell* 56, 487-493 (1989).
13. Ashkenazi, A., Ramachandran, J., and Capon, D. Acetylcholine analogue stimulates DNA synthesis in brain-derived cells via specific muscarinic acetylcholine receptor subtypes. *Nature* 340, 146-150 (1989).
14. Lammare, D., Ashkenazi, A., Fleury, S., Smith, D., Sekaly, R., and Capon, D. The MHC-binding and gp120-binding domains of CD4 are distinct and separable. *Science* 245, 743-745 (1989).
15. Ashkenazi, A., Presta, L., Marsters, S., Camerato, T., Rosenthal, K., Fendly, B., and Capon, D. Mapping the CD4 binding site for human immunodeficiency virus type 1 by alanine-scanning mutagenesis. *Proc. Natl. Acad. Sci. USA.* 87, 7150-7154 (1990).
16. Chamow, S., Peers, D., Byrn, R., Mulkerrin, M., Harris, R., Wang, W., Bjorkman, P., Capon, D., and Ashkenazi, A. Enzymatic cleavage of a CD4 immunoadhesin generates crystallizable, biologically active Fd-like fragments. *Biochemistry* 29, 9885-9891 (1990).
17. Ashkenazi, A., Smith, D., Marsters, S., Riddle, L., Gregory, T., Ho, D., and Capon, D. Resistance of primary isolates of human immunodeficiency virus type 1 to soluble CD4 is independent of CD4-gp120 binding affinity. *Proc. Natl. Acad. Sci. USA.* 88, 7056-7060 (1991).
18. Ashkenazi, A., Marsters, S., Capon, D., Chamow, S., Figari, I., Pennica, D., Goeddel, D., Palladino, M., and Smith, D. Protection against endotoxic shock by a tumor necrosis factor receptor immunoadhesin. *Proc. Natl. Acad. Sci. USA.* 88, 10535-10539 (1991).
19. Moore, J., McKeating, J., Huang, Y., Ashkenazi, A., and Ho, D. Virions of primary HIV-1 isolates resistant to sCD4 neutralization differ in sCD4 affinity and glycoprotein gp120 retention from sCD4-sensitive isolates. *J. Virol.* 66, 235-243 (1992).

20. Jin, H., Oksenberg, D., Ashkenazi, A., Peroutka, S., Duncan, A., Rozmahel, R., Yang, Y., Mengod, G., Palacios, J., and O'Dowd, B. Characterization of the human 5-hydroxytryptamine<sub>1B</sub> receptor. *J. Biol. Chem.* 267, 5735-5738 (1992).
21. Marsters, A., Frutkin, A., Simpson, N., Fendly, B. and Ashkenazi, A. Identification of cysteine-rich domains of the type 1 tumor necrosis receptor involved in ligand binding. *J. Biol. Chem.* 267, 5747-5750 (1992).
22. Chamow, S., Kogan, T., Peers, D., Hastings, R., Byrn, R., and Ashkenazi, A. Conjugation of sCD4 without loss of biological activity via a novel carbohydrate-directed cross-linking reagent. *J. Biol. Chem.* 267, 15916-15922 (1992).
23. Oksenberg, D., Marsters, A., O'Dowd, B., Jin, H., Havlik, S., Peroutka, S., and Ashkenazi, A. A single amino-acid difference confers major pharmacologic variation between human and rodent 5-HT<sub>1B</sub> receptors. *Nature* 360, 161-163 (1992).
24. Haak-Frendscho, M., Marsters, S., Chamow, S., Peers, D., Simpson, N., and Ashkenazi, A. Inhibition of interferon  $\gamma$  by an interferon  $\gamma$  receptor immunoadhesin. *Immunology* 79, 594-599 (1993).
25. Penica, D., Lam, V., Weber, R., Kohr, W., Basa, L., Spellman, M., Ashkenazi, A., Shire, S., and Goeddel, D. Biochemical characterization of the extracellular domain of the 75-kd tumor necrosis factor receptor. *Biochemistry* 32, 3131-3138. (1993).
26. Barfod, L., Zheng, Y., Kuang, W., Hart, M., Evans, T., Cerione, R., and Ashkenazi, A. Cloning and expression of a human CDC42 GTPase Activating Protein reveals a functional SH3-binding domain. *J. Biol. Chem.* 268, 26059-26062 (1993).
27. Chamow, S., Zhang, D., Tan, X., Mhtre, S., Marsters, S., Peers, D., Byrn, R., Ashkenazi, A., and Yunghans, R. A humanized bispecific immunoadhesin-antibody that retargets CD3<sup>+</sup> effectors to kill HIV-1-infected cells. *J. Immunol.* 153, 4268-4280 (1994).
28. Means, R., Krantz, S., Luna, J., Marsters, S., and Ashkenazi, A. Inhibition of murine erythroid colony formation in vitro by interferon  $\gamma$  and correction by interferon  $\gamma$  receptor immunoadhesin. *Blood* 83, 911-915 (1994).
29. Haak-Frendscho, M., Marsters, S., Mordenti, J., Gillet, N., Chen, S., and Ashkenazi, A. Inhibition of TNF by a TNF receptor immunoadhesin: comparison with an anti-TNF mAb. *J. Immunol.* 152, 1347-1353 (1994).

30. Chamow, S., Kogan, T., Venuti, M., Gadek, T., Peers, D., Mordenti, J., Shak, S., and Ashkenazi, A. Modification of CD4 immunoadhesin with monomethoxy-PEG aldehyde via reductive alkylation. *Bioconj. Chem.* 5, 133-140 (1994).
31. Jin, H., Yang, R., Marsters, S., Bunting, S., Wurm, F., Chamow, S., and Ashkenazi, A. Protection against rat endotoxic shock by p55 tumor necrosis factor (TNF) receptor immunoadhesin: comparison to anti-TNF monoclonal antibody. *J. Infect. Diseases* 170, 1323-1326 (1994).
32. Beck, J., Marsters, S., Harris, R., Ashkenazi, A., and Chamow, S. Generation of soluble interleukin-1 receptor from an immunoadhesin by specific cleavage. *Mol. Immunol.* 31, 1335-1344 (1994).
33. Pitti, B., Marsters, M., Haak-Frendscho, M., Osaka, G., Mordenti, J., Chamow, S., and Ashkenazi, A. Molecular and biological properties of an interleukin-1 receptor immunoadhesin. *Mol. Immunol.* 31, 1345-1351 (1994).
34. Oksenberg, D., Havlik, S., Peroutka, S., and Ashkenazi, A. The third intracellular loop of the 5-HT<sub>2</sub> receptor specifies effector coupling. *J. Neurochem.* 64, 1440-1447 (1995).
35. Bach, E., Szabo, S., Dighe, A., Ashkenazi, A., Aguet, M., Murphy, K., and Schreiber, R. Ligand-induced autoregulation of IFN- $\gamma$  receptor  $\beta$  chain expression in T helper cell subsets. *Science* 270, 1215-1218 (1995).
36. Jin, H., Yang, R., Marsters, S., Ashkenazi, A., Bunting, S., Marra, M., Scott, R., and Baker, J. Protection against endotoxic shock by bactericidal/permeability-increasing protein in rats. *J. Clin. Invest.* 95, 1947-1952 (1995).
37. Marsters, S., Penica, D., Bach, E., Schreiber, R., and Ashkenazi, A. Interferon  $\gamma$  signals via a high-affinity multisubunit receptor complex that contains two types of polypeptide chain. *Proc. Natl. Acad. Sci. USA.* 92, 5401-5405 (1995).
38. Van Zee, K., Moldawer, L., Oldenburg, H., Thompson, W., Stackpole, S., Montegut, W., Rogy, M., Meschter, C., Gallati, H., Schiller, C., Richter, W., Loetcher, H., Ashkenazi, A., Chamow, S., Wurm, F., Calvano, S., Lowry, S., and Lesslauer, W. Protection against lethal *E. coli* bacteremia in baboons by pretreatment with a 55-kDa TNF receptor-Ig fusion protein, Ro45-2081. *J. Immunol.* 156, 2221-2230 (1996).
39. Pitti, R., Marsters, S., Ruppert, S., Donahue, C., Moore, A., and Ashkenazi, A. Induction of apoptosis by Apo-2 Ligand, a new member of the tumor necrosis factor cytokine family. *J. Biol. Chem.* 271, 12687-12690 (1996).



40. Marsters, S., Pitti, R., Donahue, C., Rupert, S., Bauer, K., and Ashkenazi, A. Activation of apoptosis by Apo-2 ligand is independent of FADD but blocked by CrmA. *Curr. Biol.* 6, 1669-1676 (1996).
41. Marsters, S., Skubatch, M., Gray, C., and Ashkenazi, A. Herpesvirus entry mediator, a novel member of the tumor necrosis factor receptor family, activates the NF- $\kappa$ B and AP-1 transcription factors. *J. Biol. Chem.* 272, 14029-14032 (1997).
42. Sheridan, J., Marsters, S., Pitti, R., Gurney, A., Skubatch, M., Baldwin, D., Ramakrishnan, L., Gray, C., Baker, K., Wood, W.I., Goddard, A., Godowski, P., and Ashkenazi, A. Control of TRAIL-induced apoptosis by a family of signaling and decoy receptors. *Science* 277, 818-821 (1997).
43. Marsters, S., Sheridan, J., Pitti, R., Gurney, A., Skubatch, M., Baldwin, D., Huang, A., Yuan, J., Goddard, A., Godowski, P., and Ashkenazi, A. A novel receptor for Apo2L/TRAIL contains a truncated death domain. *Curr. Biol.* 7, 1003-1006 (1997).
44. Marsters, A., Sheridan, J., Pitti, R., Brush, J., Goddard, A., and Ashkenazi, A. Identification of a ligand for the death-domain-containing receptor Apo3. *Curr. Biol.* 8, 525-528 (1998).
45. Rieger, J., Naumann, U., Glaser, T., Ashkenazi, A., and Weller, M. Apo2 ligand: a novel weapon against malignant glioma? *FEBS Lett.* 427, 124-128 (1998).
46. Pender, S., Fell, J., Chamow, S., Ashkenazi, A., and MacDonald, T. A p55 TNF receptor immunoadhesin prevents T cell mediated intestinal injury by inhibiting matrix metalloproteinase production. *J. Immunol.* 160, 4098-4103 (1998).
47. Pitti, R., Marsters, S., Lawrence, D., Roy, Kischkel, F., M., Dowd, P., Huang, A., Donahue, C., Sherwood, S., Baldwin, D., Godowski, P., Wood, W., Gurney, A., Hillan, K., Cohen, R., Goddard, A., Botstein, D., and Ashkenazi, A. Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer. *Nature* 396, 699-703 (1998).
48. Mori, S., Marakami-Mori, K., Nakamura, S., Ashkenazi, A., and Bonavida, B. Sensitization of AIDS Kaposi's sarcoma cells to Apo-2 ligand-induced apoptosis by actinomycin D. *J. Immunol.* 162, 5616-5623 (1999).
49. Gurney, A., Marsters, S., Huang, A., Pitti, R., Mark, M., Baldwin, D., Gray, A., Dowd, P., Brush, J., Heldens, S., Schow, P., Goddard, A., Wood, W., Baker, K., Godowski, P., and Ashkenazi, A. Identification of a new member of the tumor necrosis factor family and its receptor, a human ortholog of mouse GITR. *Curr. Biol.* 9, 215-218 (1999).

50. Ashkenazi, A., Pai, R., Fong, s., Leung, S., Lawrence, D., Marsters, S., Blackie, C., Chang, L., McMurtrey, A., Hebert, A., DeForge, L., Khoumenis, I., Lewis, D., Harris, L., Bussiere, J., Koeppen, H., Shahrokh, Z., and Schwall, R. Safety and anti-tumor activity of recombinant soluble Apo2 ligand. *J. Clin. Invest.* **104**, 155-162 (1999).
51. Chuntharapai, A., Gibbs, V., Lu, J., Ow, A., Marsters, S., Ashkenazi, A., De Vos, A., Kim, K.J. Determination of residues involved in ligand binding and signal transmissiion in the human IFN- $\alpha$  receptor 2. *J. Immunol.* **163**, 766-773 (1999).
52. Johnsen, A.-C., Haux, J., Steinkjer, B., Nonstad, U., Egeberg, K., Sundan, A., Ashkenazi, A., and Espevik, T. Regulation of Apo2L/TRAIL expression in NK cells – involvement in NK cell-mediated cytotoxicity. *Cytokine* **11**, 664-672 (1999).
53. Roth, W., Isenmann, S., Naumann, U., Kugler, S., Bahr, M., Dichgans, J., Ashkenazi, A., and Weller, M. Eradication of intracranial human malignant glioma xenografts by Apo2L/TRAIL. *Biochem. Biophys. Res. Commun.* **265**, 479-483 (1999).
54. Hymowitz, S.G., Christinger, H.W., Fuh, G., Ultsch, M., O'Connell, M., Kelley, R.F., Ashkenazi, A. and de Vos, A.M. Triggering Cell Death: The Crystal Structure of Apo2L/TRAIL in a Complex with Death Receptor 5. *Molec. Cell* **4**, 563–571 (1999).
55. Hymowitz, S.G., O'Connel, M.P., Utsch, M.H., Hurst, A., Totpal, K., Ashkenazi, A., de Vos, A.M., Kelley, R.F. A unique zinc-binding site revealed by a high-resolution X-ray structure of homotrimeric Apo2L/TRAIL. *Biochemistry* **39**, 633-640 (2000).
56. Zhou, Q., Fukushima, P., DeGraff, W., Mitchell, J.B., Stetler-Stevenson, M., Ashkenazi, A., and Steeg, P.S. Radiation and the Apo2L/TRAIL apoptotic pathway preferentially inhibit the colonization of premalignant human breast cancer cells overexpressing cyclin D1. *Cancer Res.* **60**, 2611-2615 (2000).
57. Kischkel, F.C., Lawrence, D. A., Chuntharapai, A., Schow, P., Kim, J., and Ashkenazi, A. Apo2L/TRAIL-dependent recruitment of endogenous FADD and Caspase-8 to death receptors 4 and 5. *Immunity* **12**, 611-620 (2000).
58. Yan, M., Marsters, S.A., Grewal, I.S., Wang, H., \*Ashkenazi, A., and \*Dixit, V.M. Identification of a receptor for BlyS demonstrates a crucial role in humoral immunity. *Nature Immunol.* **1**, 37-41 (2000).

59. Marsters, S.A., Yan, M., Pitti, R.M., Haas, P.E., Dixit, V.M., and Ashkenazi, A. Interaction of the TNF homologues BLyS and APRIL with the TNF receptor homologues BCMA and TACI. *Curr. Biol.* 10, 785-788 (2000).
60. Kischkel, F.C., and Ashkenazi, A. Combining enhanced metabolic labeling with immunoblotting to detect interactions of endogenous cellular proteins. *Biotechniques* 29, 506-512 (2000).
61. Lawrence, D., Shahrokh, Z., Marsters, S., Achilles, K., Shih, D., Mounho, B., Hillan, K., Totpal, K., DeForge, L., Schow, P., Hooley, J., Sherwood, S., Pai, R., Leung, S., Khan, L., Gliniak, B., Bussiere, J., Smith, C., Strom, S., Kelley, S., Fox, J., Thomas, D., and Ashkenazi, A. Differential hepatocyte toxicity of recombinant Apo2L/TRAIL versions. *Nature Med.* 7, 383-385 (2001).
62. Chuntharapai, A., Dodge, K., Grimmer, K., Schroeder, K., Marsters, S.A., Koeppen, H., Ashkenazi, A., and Kim, K.J. Isotype-dependent inhibition of tumor growth in vivo by monoclonal antibodies to death receptor 4. *J. Immunol.* 166, 4891-4898 (2001).
63. Pollack, I.F., Erff, M., and Ashkenazi, A. Direct stimulation of apoptotic signaling by soluble Apo2L/tumor necrosis factor-related apoptosis-inducing ligand leads to selective killing of glioma cells. *Clin. Cancer Res.* 7, 1362-1369 (2001).
64. Wang, H., Marsters, S.A., Baker, T., Chan, B., Lee, W.P., Fu, L., Tumas, D., Yan, M., Dixit, V.M., \*Ashkenazi, A., and \*Grewal, I.S. TACI-ligand interactions are required for T cell activation and collagen-induced arthritis in mice. *Nature Immunol.* 2, 632-637 (2001).
65. Kischkel, F.C., Lawrence, D. A., Tinel, A., Virmani, A., Schow, P., Gazdar, A., Blenis, J., Arnott, D., and Ashkenazi, A. Death receptor recruitment of endogenous caspase-10 and apoptosis initiation in the absence of caspase-8. *J. Biol. Chem.* 276, 46639-46646 (2001).
66. LeBlanc, H., Lawrence, D.A., Varfolomeev, E., Totpal, K., Morlan, J., Schow, P., Fong, S., Schwall, R., Sinicropi, D., and Ashkenazi, A. Tumor cell resistance to death receptor induced apoptosis through mutational inactivation of the proapoptotic Bcl-2 homolog Bax. *Nature Med.* 8, 274-281 (2002).
67. Miller, K., Meng, G., Liu, J., Hurst, A., Hsei, V., Wong, W-L., Ekert, R., Lawrence, D., Sherwood, S., DeForge, L., Gaudreault, Keller, G., Sliwkowski, M., Ashkenazi, A., and Presta, L. Design, Construction, and analyses of multivalent antibodies. *J. Immunol.* 170, 4854-4861 (2003).

68. Varfolomeev, E., Kischkel, F., Martin, F., Wanh, H., Lawrence, D., Olsson, C., Tom, L., Erickson, S., French, D., Schow, P., Grewal, I. and Ashkenazi, A. Immune system development in APRIL knockout mice. Submitted.

**Review articles:**

1. Ashkenazi, A., Peralta, E., Winslow, J., Ramachandran, J., and Capon, D., J. Functional role of muscarinic acetylcholine receptor subtype diversity. *Cold Spring Harbor Symposium on Quantitative Biology*. **LIII**, 263-272 (1988).
2. Ashkenazi, A., Peralta, E., Winslow, J., Ramachandran, J., and Capon, D. Functional diversity of muscarinic receptor subtypes in cellular signal transduction and growth. *Trends Pharmacol. Sci.* Dec Supplement, 12-21 (1989).
3. Chamow, S., Duliège, A., Ammann, A., Kahn, J., Allen, D., Eichberg, J., Byrn, R., Capon, D., Ward, R., and Ashkenazi, A. CD4 immunoadhesins in anti-HIV therapy: new developments. *Int. J. Cancer* Supplement 7, 69-72 (1992).
4. Ashkenazi, A., Capon, and D. Ward, R. Immunoadhesins. *Int. Rev. Immunol.* **10**, 217-225 (1993).
5. Ashkenazi, A., and Peralta, E. Muscarinic Receptors. In *Handbook of Receptors and Channels*. (S. Peroutka, ed.), CRC Press, Boca Raton, Vol. I, p. 1-27, (1994).
6. Krantz, S. B., Means, R. T., Jr., Lina, J., Marsters, S. A., and Ashkenazi, A. Inhibition of erythroid colony formation in vitro by gamma interferon. In *Molecular Biology of Hematopoiesis* (N. Abraham, R. Shadduck, A. Levine F. Takaku, eds.) Intercept Ltd. Paris, Vol. 3, p. 135-147 (1994).
7. Ashkenazi, A. Cytokine neutralization as a potential therapeutic approach for SIRS and shock. *J. Biotechnology in Healthcare* **1**, 197-206 (1994).
8. Ashkenazi, A., and Chamow, S. M. Immunoadhesins: an alternative to human monoclonal antibodies. *Immunomethods: A companion to Methods in Enzimology* **8**, 104-115 (1995).
9. Chamow, S., and Ashkenazi, A. Immunoadhesins: Principles and Applications. *Trends Biotech.* **14**, 52-60 (1996).
10. Ashkenazi, A., and Chamow, S. M. Immunoadhesins as research tools and therapeutic agents. *Curr. Opin. Immunol.* **9**, 195-200 (1997).
11. Ashkenazi, A., and Dixit, V. Death receptors: signaling and modulation. *Science* **281**, 1305-1308 (1998).
12. Ashkenazi, A., and Dixit, V. Apoptosis control by death and decoy receptors. *Curr. Opin. Cell. Biol.* **11**, 255-260 (1999).

13. Ashkenazi, A. Chapters on Apo2L/TRAIL; DR4, DR5, DcR1, DcR2; and DcR3. Online Cytokine Handbook ([www.apnet.com/cytokinereference/](http://www.apnet.com/cytokinereference/)).
14. Ashkenazi, A. Targeting death and decoy receptors of the tumor necrosis factor superfamily. *Nature Rev. Cancer* 2, 420-430 (2002).
15. LeBlanc, H. and Ashkenazi, A. Apoptosis signaling by Apo2L/TRAIL. *Cell Death and Differentiation* 10, 66-75 (2003).
16. Almasan, A. and Ashkenazi, A. Apo2L/TRAIL: apoptosis signaling, biology, and potential for cancer therapy. *Cytokine and Growth Factor Reviews* 14, 337-348 (2003).

**Book:**

Antibody Fusion Proteins (Chamow, S., and Ashkenazi, A., eds., John Wiley and Sons Inc.) (1999).

**Talks:**

1. Resistance of primary HIV isolates to CD4 is independent of CD4-gp120 binding affinity. UCSD Symposium, HIV Disease: Pathogenesis and Therapy. Greenelefe, FL, March 1991.
2. Use of immuno-hybrids to extend the half-life of receptors. IBC conference on Biopharmaceutical Half-life Extension. New Orleans, LA, June 1992.
3. Results with TNF receptor Immunoadhesins for the Treatment of Sepsis. IBC conference on Endotoxemia and Sepsis. Philadelphia, PA, June 1992.
4. Immunoadhesins: an alternative to human antibodies. IBC conference on Antibody Engineering. San Diego, CA, December 1993.
5. Tumor necrosis factor receptor: a potential therapeutic for human septic shock. American Society for Microbiology Meeting, Atlanta, GA, May 1993.
6. Protective efficacy of TNF receptor immunoadhesin vs anti-TNF monoclonal antibody in a rat model for endotoxic shock. 5th International Congress on TNF. Asilomar, CA, May 1994.
7. Interferon- $\gamma$  signals via a multisubunit receptor complex that contains two types of polypeptide chain. American Association of Immunologists Conference. San Francisco, CA, July 1995.
8. Immunoadhesins: Principles and Applications. Gordon Research Conference on Drug Delivery in Biology and Medicine. Ventura, CA, February 1996.

9. Apo-2 Ligand, a new member of the TNF family that induces apoptosis in tumor cells. Cambridge Symposium on TNF and Related Cytokines in Treatment of Cancer. Hilton-Head, NC, March 1996.
10. Induction of apoptosis by Apo2 Ligand. American Society for Biochemistry and Molecular Biology, Symposium on Growth Factors and Cytokine Receptors. New Orleans, LA, June, 1996.
11. Apo2 ligand, an extracellular trigger of apoptosis. 2nd Clontech Symposium, Palo Alto, CA, October 1996.
12. Regulation of apoptosis by members of the TNF ligand and receptor families. Stanford University School of Medicine, Palo Alto, CA, December 1996.
13. Apo-3: a novel receptor that regulates cell death and inflammation. 4th International Congress on Immune Consequences of Trauma, Shock, and Sepsis. Munich, Germany, March 1997.
14. New members of the TNF ligand and receptor families that regulate apoptosis, inflammation, and immunity. UCLA School of Medicine, LA, CA, March 1997.
15. Immunoadhesins: an alternative to monoclonal antibodies. 5th World Conference on Bispecific Antibodies. Volendam, Holland, June 1997.
16. Control of Apo2L signaling. Cold Spring Harbor Laboratory Symposium on Programmed Cell Death. Cold Spring Harbor, New York. September, 1997.
17. Chairman and speaker, Apoptosis Signaling session. IBC's 4th Annual Conference on Apoptosis. San Diego, CA., October 1997.
18. Control of Apo2L signaling by death and decoy receptors. American Association for the Advancement of Science. Philadelphia, PA, February 1998.
19. Apo2 ligand and its receptors. American Society of Immunologists. San Francisco, CA, April 1998.
20. Death receptors and ligands. 7th International TNF Congress. Cape Cod, MA, May 1998.
21. Apo2L as a potential therapeutic for cancer. UCLA School of Medicine. LA, CA, June 1998.
22. Apo2L as a potential therapeutic for cancer. Gordon Research Conference on Cancer Chemotherapy. New London, NH, July 1998.
23. Control of apoptosis by Apo2L. Endocrine Society Conference, Stevenson, WA, August 1998.
24. Control of apoptosis by Apo2L. International Cytokine Society Conference, Jerusalem, Israel, October 1998.

25. Apoptosis control by death and decoy receptors. American Association for Cancer Research Conference, Whistler, BC, Canada, March 1999.
26. Apoptosis control by death and decoy receptors. American Society for Biochemistry and Molecular Biology Conference, San Francisco, CA, May 1999.
27. Apoptosis control by death and decoy receptors. Gordon Research Conference on Apoptosis, New London, NH, June 1999.
28. Apoptosis control by death and decoy receptors. Arthritis Foundation Research Conference, Alexandria GA, Aug 1999.
29. Safety and anti-tumor activity of recombinant soluble Apo2L/TRAIL. Cold Spring Harbor Laboratory Symposium on Programmed Cell Death. . Cold Spring Harbor, NY, September 1999.
30. The Apo2L/TRAIL system: therapeutic potential. American Association for Cancer Research, Lake Tahoe, NV, Feb 2000.
31. Apoptosis and cancer therapy. Stanford University School of Medicine, Stanford, CA, Mar 2000.
32. Apoptosis and cancer therapy. University of Pennsylvania School of Medicine, Philadelphia, PA, Apr 2000.
33. Apoptosis signaling by Apo2L/TRAIL. International Congress on TNF. Trondheim, Norway, May 2000.
34. The Apo2L/TRAIL system: therapeutic potential. Cap-CURE summit meeting. Santa Monica, CA, June 2000.
35. The Apo2L/TRAIL system: therapeutic potential. MD Anderson Cancer Center. Houston, TX, June 2000.
36. Apoptosis signaling by Apo2L/TRAIL. The Protein Society, 14<sup>th</sup> Symposium. San Diego, CA, August 2000.
37. Anti-tumor activity of Apo2L/TRAIL. AAPS annual meeting. Indianapolis, IN Aug 2000.
38. Apoptosis signaling and anti-cancer potential of Apo2L/TRAIL. Cancer Research Institute, UC San Francisco, CA, September 2000.
39. Apoptosis signaling by Apo2L/TRAIL. Kenote address, TNF family Minisymposium, NIH. Bethesda, MD, September 2000.
40. Death receptors: signaling and modulation. Keystone symposium on the Molecular basis of cancer. Taos, NM, Jan 2001.
41. Preclinical studies of Apo2L/TRAIL in cancer. Symposium on Targeted therapies in the treatment of lung cancer. Aspen, CO, Jan 2001.

42. Apoptosis signaling by Apo2L/TRAIL. Weizmann Institute of Science, Rehovot, Israel, March 2001.
43. Apo2L/TRAIL: Apoptosis signaling and potential for cancer therapy. Weizmann Institute of Science, Rehovot, Israel, March 2001.
44. Targeting death receptors in cancer with Apo2L/TRAIL. Cell Death and Disease conference, North Falmouth, MA, Jun 2001.
45. Targeting death receptors in cancer with Apo2L/TRAIL. Biotechnology Organization conference, San Diego, CA, Jun 2001.
46. Apo2L/TRAIL signaling and apoptosis resistance mechanisms. Gordon Research Conference on Apoptosis, Oxford, UK, July 2001.
47. Apo2L/TRAIL signaling and apoptosis resistance mechanisms. Cleveland Clinic Foundation, Cleveland, OH, Oct 2001.
48. Apoptosis signaling by death receptors: overview. International Society for Interferon and Cytokine Research conference, Cleveland, OH, Oct 2001.
49. Apoptosis signaling by death receptors. American Society of Nephrology Conference. San Francisco, CA, Oct 2001.
50. Targeting death receptors in cancer. Apoptosis: commercial opportunities. San Diego, CA, Apr 2002.
51. Apo2L/TRAIL signaling and apoptosis resistance mechanisms. Kimmel Cancer Research Center, Johns Hopkins University, Baltimore MD. May 2002.
52. Apoptosis control by Apo2L/TRAIL. (Keynote Address) University of Alabama Cancer Center Retreat, Birmingham, Ab. October 2002.
53. Apoptosis signaling by Apo2L/TRAIL. (Session co-chair) TNF international conference. San Diego, CA. October 2002.
54. Apoptosis signaling by Apo2L/TRAIL. Swiss Institute for Cancer Research (ISREC). Lausanne, Swizerland. Jan 2003.
55. Apoptosis induction with Apo2L/TRAIL. Conference on New Targets and Innovative Strategies in Cancer Treatment. Monte Carlo. February 2003.
56. Apoptosis signaling by Apo2L/TRAIL. Hermelin Brain Tumor Center Symposium on Apoptosis. Detroit, MI. April 2003.
57. Targeting apoptosis through death receptors. Sixth Annual Conference on Targeted Therapies in the Treatment of Breast Cancer. Kona, Hawaii. July 2003.
58. Targeting apoptosis through death receptors. Second International Conference on Targeted Cancer Therapy. Washington, DC. Aug 2003.

**Issued Patents:**



1. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 5,329,028 (Jul 12, 1994).
2. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 5,605,791 (Feb 25, 1997).
3. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 5,889,155 (Jul 27, 1999).
4. Ashkenazi, A., APO-2 Ligand. US patent 6,030,945 (Feb 29, 2000).
5. Ashkenazi, A., Chuntharapai, A., Kim, J., APO-2 ligand antibodies. US patent 6,046,048 (Apr 4, 2000).
6. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 6,124,435 (Sep 26, 2000).
7. Ashkenazi, A., Chuntharapai, A., Kim, J., Method for making monoclonal and cross-reactive antibodies. US patent 6,252,050 (Jun 26, 2001).
8. Ashkenazi, A. APO-2 Receptor. US patent 6,342,369 (Jan 29, 2002).
9. Ashkenazi, A. Fong, S., Goddard, A., Gurney, A., Napier, M., Tumas, D., Wood, W. A-33 polypeptides. US patent 6,410,708 (Jun 25, 2002).
10. Ashkenazi, A. APO-3 Receptor. US patent 6,462,176 B1 (Oct 8, 2002).
11. Ashkenazi, A. APO-2LI and APO-3 polypeptide antibodies. US patent 6,469,144 B1 (Oct 22, 2002).
12. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 6,582,928B1 (Jun 24, 2003).

## DECLARATION OF PAUL POLAKIS, Ph.D.

I, Paul Polakis, Ph.D., declare and say as follows:

1. I was awarded a Ph.D. by the Department of Biochemistry of the Michigan State University in 1984. My scientific Curriculum Vitae is attached to and forms part of this Declaration (Exhibit A).
2. I am currently employed by Genentech, Inc. where my job title is Staff Scientist. Since joining Genentech in 1999, one of my primary responsibilities has been leading Genentech's Tumor Antigen Project, which is a large research project with a primary focus on identifying tumor cell markers that find use as targets for both the diagnosis and treatment of cancer in humans.
3. As part of the Tumor Antigen Project, my laboratory has been analyzing differential expression of various genes in tumor cells relative to normal cells. The purpose of this research is to identify proteins that are abundantly expressed on certain tumor cells and that are either (i) not expressed, or (ii) expressed at lower levels, on corresponding normal cells. We call such differentially expressed proteins "tumor antigen proteins". When such a tumor antigen protein is identified, one can produce an antibody that recognizes and binds to that protein. Such an antibody finds use in the diagnosis of human cancer and may ultimately serve as an effective therapeutic in the treatment of human cancer.
4. In the course of the research conducted by Genentech's Tumor Antigen Project, we have employed a variety of scientific techniques for detecting and studying differential gene expression in human tumor cells relative to normal cells, at genomic DNA, mRNA and protein levels. An important example of one such technique is the well known and widely used technique of microarray analysis which has proven to be extremely useful for the identification of mRNA molecules that are differentially expressed in one tissue or cell type relative to another. In the course of our research using microarray analysis, we have identified approximately 200 gene transcripts that are present in human tumor cells at significantly higher levels than in corresponding normal human cells. To date, we have generated antibodies that bind to about 30 of the tumor antigen proteins expressed from these differentially expressed gene transcripts and have used these antibodies to quantitatively determine the level of production of these tumor antigen proteins in both human cancer cells and corresponding normal cells. We have then compared the levels of mRNA and protein in both the tumor and normal cells analyzed.
5. From the mRNA and protein expression analyses described in paragraph 4 above, we have observed that there is a strong correlation between changes in the level of mRNA present in any particular cell type and the level of protein

expressed from that mRNA in that cell type. In approximately 80% of our observations we have found that increases in the level of a particular mRNA correlates with changes in the level of protein expressed from that mRNA when human tumor cells are compared with their corresponding normal cells.

6. Based upon my own experience accumulated in more than 20 years of research, including the data discussed in paragraphs 4 and 5 above and my knowledge of the relevant scientific literature, it is my considered scientific opinion that for human genes, an increased level of mRNA in a tumor cell relative to a normal cell typically correlates to a similar increase in abundance of the encoded protein in the tumor cell relative to the normal cell. In fact, it remains a central dogma in molecular biology that increased mRNA levels are predictive of corresponding increased levels of the encoded protein. While there have been published reports of genes for which such a correlation does not exist, it is my opinion that such reports are exceptions to the commonly understood general rule that increased mRNA levels are predictive of corresponding increased levels of the encoded protein.

7. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

Dated: 5/07/04

By: Paul Polakis

Paul Polakis, Ph.D.

## CURRICULUM VITAE

PAUL G. POLAKIS  
Staff Scientist  
Genentech, Inc  
1 DNA Way, MS#40  
S. San Francisco, CA 94080

### EDUCATION:

Ph.D., Biochemistry, Department of Biochemistry,  
Michigan State University (1984)

B.S., Biology. College of Natural Science, Michigan State University (1977)

### PROFESSIONAL EXPERIENCE:

2002-present	Staff Scientist, Genentech, Inc S. San Francisco, CA
1999- 2002	Senior Scientist, Genentech, Inc., S. San Francisco, CA
1997 -1999	Research Director Onyx Pharmaceuticals, Richmond, CA
1992- 1996	Senior Scientist, Project Leader, Onyx Pharmaceuticals, Richmond, CA
<hr/>	
1991-1992	Senior Scientist, Chiron Corporation, Emeryville, CA.
1989-1991	Scientist, Cetus Corporation, Emeryville CA.
1987-1989	Postdoctoral Research Associate, Genentech, Inc., South San Francisco, CA.
1985-1987	Postdoctoral Research Associate, Department of Medicine, Duke University Medical Center, Durham, NC

1984-1985

Assistant Professor, Department of Chemistry,  
Oberlin College, Oberlin, Ohio

1980-1984

Graduate Research Assistant, Department of  
Biochemistry, Michigan State University  
East Lansing, Michigan

### **PUBLICATIONS:**

1. Polakis, P. G. and Wilson, J. E. 1982 Purification of a Highly Bindable Rat Brain Hexokinase by High Performance Liquid Chromatography. **Biochem. Biophys. Res. Commun.** 107, 937-943.

2. Polakis, P.G. and Wilson, J. E. 1984 Proteolytic Dissection of Rat Brain Hexokinase: Determination of the Cleavage Pattern during Limited Digestion with Trypsin. **Arch. Biochem. Biophys.** 234, 341-352.

3. Polakis, P. G. and Wilson, J. E. 1985 An Intact Hydrophobic N-Terminal Sequence is Required for the Binding Rat Brain Hexokinase to Mitochondria. **Arch. Biochem. Biophys.** 236, 328-337.

4. Uhing, R.J., Polakis, P.G. and Snyderman, R. 1987 Isolation of GTP-binding Proteins from Myeloid HL60 Cells. **J. Biol. Chem.** 262, 15575-15579.

5. Polakis, P.G., Uhing, R.J. and Snyderman, R. 1988 The Formylpeptide Chemoattractant Receptor Copurifies with a GTP-binding Protein Containing a Distinct 40 kDa Pertussis Toxin Substrate. **J. Biol. Chem.** 263, 4969-4979.

6. Uhing, R. J., Dillon, S., Polakis, P. G., Truett, A. P. and Snyderman, R. 1988 Chemoattractant Receptors and Signal Transduction Processes in Cellular and Molecular Aspects of Inflammation ( Poste, G. and Crooke, S. T. eds.) pp 335-379.

7. Polakis, P.G., Evans, T. and Snyderman 1989 Multiple Chromatographic Forms of the Formylpeptide Chemoattractant Receptor and their Relationship to GTP-binding Proteins. **Biochem. Biophys. Res. Commun.** 161, 276-283.

8. Polakis, P. G., Snyderman, R. and Evans, T. 1989 Characterization of G25K, a GTP-binding Protein Containing a Novel Putative Nucleotide Binding Domain. **Biochem. Biophys. Res. Commun.** 160, 25-32.

9. Polakis, P., Weber, R.F., Nevins, B., Didsbury, J. Evans, T. and Snyderman, R. 1989 Identification of the *ral* and *rac1* Gene Products, Low Molecular Mass GTP-binding Proteins from Human Platelets. **J. Biol. Chem.** 264, 16383-16389.

10. Snyderman, R., Perianin, A., Evans, T., Polakis, P. and Didsbury, J. 1989 G Proteins and Neutrophil Function. In ADP-Ribosylating Toxins and G Proteins: Insights into Signal Transduction. ( J. Moss and M. Vaughn, eds.) Amer. Soc. Microbiol. pp. 295-323.

11. Hart, M.J., Polakis, P.G., Evans, T. and Cerrione, R.A. 1990 The identification and Characterization of an Epidermal Growth Factor-Stimulated Phosphorylation of a Specific Low Molecular Mass GTP-binding Protein in a Reconstituted Phospholipid Vesicle System. *J. Biol. Chem.* 265, 5990-6001.

12. Yatani, A., Okabe, K., Polakis, P., Halenbeck, R., McCormick, F. and Brown, A. M. 1990 ras p21 and GAP Inhibit Coupling of Muscarinic Receptors to Atrial K<sup>+</sup> Channels. *Cell*. 61, 769-776.

13. Munemitsu, S., Innis, M.A., Clark, R., McCormick, F., Ullrich, A. and Polakis, P.G. 1990 Molecular Cloning and Expression of a G25K cDNA, the Human Homolog of the Yeast Cell Cycle Gene CDC42. *Mol. Cell. Biol.* 10, 5977-5982.

14. Polakis, P.G. Rubinfeld, B. Evans, T. and McCormick, F. 1991 Purification of Plasma Membrane-Associated GTPase Activating Protein (GAP) Specific for rap-1/krev-1 from HL60 Cells. *Proc. Natl. Acad. Sci. USA* 88, 239-243.

15. Moran, M. F., Polakis, P., McCormick, F., Pawson, T. and Ellis, C. 1991 Protein Tyrosine Kinases Regulate the Phosphorylation, Protein Interactions, Subcellular Distribution, and Activity of p21ras GTPase Activating Protein. *Mol. Cell. Biol.* 11, 1804-1812

16. Rubinfeld, B., Wong, G., Bekesi, E. Wood, A. McCormick, F. and Polakis, P. G. 1991 A Synthetic Peptide Corresponding to a Sequence in the GTPase Activating Protein Inhibits p21<sup>ras</sup> Stimulation and Promotes Guanine Nucleotide Exchange. *Internatl. J. Peptide and Prot. Res.* 38, 47-53.

17. Rubinfeld, B., Munemitsu, S., Clark, R., Conroy, L., Watt, K., Crosier, W., McCormick, F., and Polakis, P. 1991 Molecular Cloning of a GTPase Activating Protein Specific for the Krev-1 Protein p21<sup>rap1</sup>. *Cell* 65, 1033-1042.

18. Zhang, K. Papageorge, A., G., Martin, P., Vass, W. C., Olah, Z., Polakis, P., McCormick, F. and Lowy, D. R. 1991 Heterogenous Amino Acids in RAS and Rap1A Specifying Sensitivity to GAP Proteins. *Science* 254, 1630-1634.

19. Martin, G., Yatani, A., Clark, R., Polakis, P., Brown, A. M. and McCormick, F. 1992 GAP Domains Responsible for p21<sup>ras</sup>-dependent Inhibition of Muscarinic Atrial K<sup>+</sup> Channel Currents. *Science* 255, 192-194.

20. McCormick, F., Martin, G. A., Clark, R., Bollag, G. and Polakis, P. 1992 Regulation of p21ras by GTPase Activating Proteins. Cold Spring Harbor **Symposia on Quantitative Biology**. Vol. 56, 237-241.

21. Pronk, G. B., Polakis, P., Wong, G., deVries-Smits, A. M., Bos J. L. and McCormick, F. 1992 p60<sup>v-src</sup> Can Associate with and Phosphorylate the p21<sup>ras</sup> GTPase Activating Protein. *Oncogene* 7, 389-394.

22. Polakis P. and McCormick, F. 1992 Interactions Between p21<sup>ras</sup> Proteins and Their GTPase Activating Proteins. In Cancer Surveys (Franks, L. M., ed.) 12, 25-42.

23. Wong, G., Muller, O., Clark, R., Conroy, L., Moran, M., Polakis, P. and McCormick, F. 1992 Molecular cloning and nucleic acid binding properties of the GAP-associated tyrosine phosphoprotein p62. *Cell* 69, 551-558.
24. Polakis, P., Rubinfeld, B. and McCormick, F. 1992 Phosphorylation of rap1GAP in vivo and by cAMP-dependent Kinase and the Cell Cycle p34<sup>cdc2</sup> Kinase in vitro. *J. Biol. Chem.* 267, 10780-10785.
25. McCabe, P.C., Haubrauck, H., Polakis, P., McCormick, F., and Innis, M. A. 1992 Functional Interactions Between p21<sup>rap1A</sup> and Components of the Budding pathway of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 12, 4084-4092.
26. Rubinfeld, B., Crosier, W.J., Albert, I., Conroy, L., Clark, R., McCormick, F. and Polakis, P. 1992 Localization of the rap1GAP Catalytic Domain and Sites of Phosphorylation by Mutational Analysis. *Mol. Cell. Biol.* 12, 4634-4642.
27. Ando, S., Kaibuchi, K., Sasaki, K., Hiraoka, T., Nishiyama, T., Mizuno, T., Asada, M., Nunoi, H., Matsuda, I., Matsuura, Y., Polakis, P., McCormick, F. and Takai, Y. 1992 Post-translational processing of rac p21s is important both for their interaction with the GDP/GTP exchange proteins and for their activation of NADPH oxidase. *J. Biol. Chem.* 267, 25709-25713.
28. Janoueix-Lerosey, I., Polakis, P., Tavitian, A. and deGunzburg, J. 1992 Regulation of the GTPase activity of the ras-related rap2 protein. *Biochem. Biophys. Res. Commun.* 189, 455-464.
29. Polakis, P. 1993 GAPs Specific for the rap1/Krev-1 Protein. in GTP-binding Proteins: the ras-superfamily. ( J.C. LaCale and F. McCormick, eds.) 445-452.
30. Polakis, P. and McCormick, F. 1993 Structural requirements for the interaction of p21<sup>ras</sup> with GAP, exchange factors, and its biological effector target. *J. Biol. Chem.* 268, 9157-9160.
31. Rubinfeld, B., Souza, B., Albert, I., Muller, O., Chamberlain, S., Masiarz, F., Munemitsu, S. and Polakis, P. 1993 Association of the APC gene product with beta- catenin. *Science* 262, 1731-1734.
32. Weiss, J., Rubinfeld, B., Polakis, P., McCormick, F., Cavenee, W. A. and Arden, K. 1993. The gene for human rap1-GTPase activating protein (rap1GAP) maps to chromosome 1p35-1p36.1. *Cytogenet. Cell Genet.* 66, 18-21.
33. Sato, K. Y., Polakis, P., Haubruck, H., Fasching, C. L., McCormick, F. and Stanbridge, E. J. 1994 Analysis of the tumor suppressor activity of the K-rev gene in human tumor cell lines. *Cancer Res.* 54, 552-559.
34. Janoueix-Lerosey, I., Fontenay, M., Tobelem, G., Tavitian, A., Polakis, P. and DeGunzburg, J. 1994 Phosphorylation of rap1GAP during the cell cycle. *Biochem. Biophys. Res. Commun.* 202, 967-975
35. Munemitsu, S., Souza, B., Mueller, O., Albert, I., Rubinfeld, B., and Polakis, P. 1994 The APC gene product associates with microtubules in vivo and affects their assembly in vitro. *Cancer Res.* 54, 3676-3681.

36. Rubinfeld, B. and Polakis, P. 1995 Purification of baculovirus produced rap1GAP. **Methods Enz.** 255,31
37. Polakis, P. 1995 Mutations in the APC gene and their implications for protein structure and function. **Current Opinions in Genetics and Development** 5, 66-71
38. Rubinfeld, B., Souza, B., Albert, I., Munemitsu, S. and Polakis P. 1995 The APC protein and E-cadherin form similar but independent complexes with  $\alpha$ -catenin,  $\beta$ -catenin and Plakoglobin. **J. Biol. Chem.** 270, 5549-5555
39. Munemitsu, S., Albert, I., Souza, B., Rubinfeld, B., and Polakis, P. 1995 Regulation of intracellular  $\beta$ -catenin levels by the APC tumor suppressor gene. **Proc. Natl. Acad. Sci.** 92, 3046-3050.
40. Lock, P., Fumagalli, S., Polakis, P. McCormick, F. and Courtneidge, S. A. 1996 The human p62 cDNA encodes Sam68 and not the rasGAP-associated p62 protein. **Cell** 84, 23-24.
41. Papkoff, J., Rubinfeld, B., Schryver, B. and Polakis, P. 1996 Wnt-1 regulates free pools of catenins and stabilizes APC-catenin complexes. **Mol. Cell. Biol.** 16, 2128-2134.
42. Rubinfeld, B., Albert, I., Porfiri, E., Fiol, C., Munemitsu, S. and Polakis, P. 1996 Binding of GSK3 $\beta$  to the APC- $\beta$ -catenin complex and regulation of complex assembly. **Science** 272, 1023-1026.
43. Munemitsu, S., Albert, I., Rubinfeld, B. and Polakis, P. 1996 Deletion of amino-terminal structure stabilizes  $\beta$ -catenin in vivo and promotes the hyperphosphorylation of the APC tumor suppressor protein. **Mol. Cell. Biol.** 16, 4088-4094.
44. Hart, M. J., Callow, M. G., Sousa, B. and Polakis P. 1996 IQGAP1, a calmodulin binding protein with a rasGAP related domain, is a potential effector for cdc42Hs. **EMBO J.** 15, 2997-3005.
45. Nathke, I. S., Adams, C. L., Polakis, P., Sellin, J. and Nelson, W. J. 1996 The adenomatous polyposis coli (APC) tumor suppressor protein is localized to plasma membrane sites involved in active epithelial cell migration. **J. Cell. Biol.** 134, 165-180.
46. Hart, M. J., Sharma, S., elMasry, N., Qui, R-G., McCabe, P., Polakis, P. and Bollag, G. 1996 Identification of a novel guanine nucleotide exchange factor for the rho GTPase. **J. Biol. Chem.** 271, 25452.
47. Thomas JE, Smith M, Rubinfeld B, Gutowski M, Beckmann RP, and Polakis P. 1996 Subcellular localization and analysis of apparent 180-kDa and 220-kDa proteins of the breast cancer susceptibility gene, BRCA1. **J. Biol. Chem.** 1996 271, 28630-28635
48. Hayashi, S., Rubinfeld, B., Souza, B., Polakis, P., Wieschaus, E., and Levine, A. 1997 A Drosophila homolog of the tumor suppressor adenomatous polyposis coli



down-regulates  $\beta$ -catenin but its zygotic expression is not essential for the regulation of armadillo. **Proc. Natl. Acad. Sci.** 94, 242-247.

49. Vleminckx, K., Rubinfeld, B., **Polakis, P.** and Gumbiner, B. 1997 The APC tumor suppressor protein induces a new axis in *Xenopus* embryos. **J. Cell. Biol.** 136, 411-420.

50. Rubinfeld, B., Robbins, P., El-Gamil, M., Albert, I., Porfiri, P. and **Polakis, P.** 1997 Stabilization of  $\beta$ -catenin by genetic defects in melanoma cell lines. **Science** 275, 1790-1792.

51. **Polakis, P.** The adenomatous polyposis coli (APC) tumor suppressor. 1997 **Biochem. Biophys. Acta**, 1332, F127-F147.

52. Rubinfeld, B., Albert, I., Porfiri, E., Munemitsu, S., and **Polakis, P.** 1997 Loss of  $\beta$ -catenin regulation by the APC tumor suppressor protein correlates with loss of structure due to common somatic mutations of the gene. **Cancer Res.** 57, 4624-4630.

53. Porfiri, E., Rubinfeld, B., Albert, I., Hovanes, K., Waterman, M., and **Polakis, P.** 1997. Induction of a  $\beta$ -catenin-LEF-1 complex by wnt-1 and transforming mutants of  $\beta$ -catenin. **Oncogene** 15, 2833-2839.

54. Thomas JE, Smith M, Tonkinson JL, Rubinfeld B, and **Polakis P.**, 1997 Induction of phosphorylation on BRCA1 during the cell cycle and after DNA damage. **Cell Growth Differ.** 8, 801-809.

55. Hart, M., de los Santos, R., Albert, I., Rubinfeld, B., and **Polakis P.**, 1998 Down regulation of  $\beta$ -catenin by human Axin and its association with the adenomatous polyposis coli (APC) tumor suppressor,  $\beta$ -catenin and glycogen synthase kinase 3 $\beta$ . **Current Biology** 8, 573-581.

56. **Polakis, P.** 1998 The oncogenic activation of  $\beta$ -catenin. **Current Opinions in Genetics and Development** 9, 15-21

57. Matt Hart, Jean-Paul Concordet, Irina Lassot, Iris Albert, Rico del los Santos, Herve Durand, Christine Perret, Bonnee Rubinfeld, Florence Margottin, Richard Benarous and **Paul Polakis.** 1999 The F-box protein  $\beta$ -TrCP associates with phosphorylated  $\beta$ -catenin and regulates its activity in the cell. **Current Biology** 9, 207-10.

58. Howard C. Crawford, Barbara M. Fingleton, Bonnee Rubinfeld, **Paul Polakis** and Lynn M. Matrisian 1999 The metalloproteinase matrilysin is a target of  $\beta$ -catenin transactivation in intestinal tumours. **Oncogene** 18, 2883-91.

59. Meng J, Glick JL, **Polakis P.**, Casey PJ. 1999 Functional interaction between Galpha(z) and Rap1GAP suggests a novel form of cellular cross-talk. **J Biol Chem.** 17, 36663-9

60. Vijayasurian Easwaran, Virginia Song, Paul Polakis and Steve Byers 1999 The ubiquitin-proteosome pathway and serine kinase activity modulate APC mediated regulation of  $\beta$ -catenin-LEF signaling. *J. Biol. Chem.* 274(23):16641-5.
61. Polakis P, Hart M and Rubinfeld B. 1999. Defects in the regulation of beta-catenin in colorectal cancer. *Adv Exp Med Biol.* 470, 23-32
62. Shen Z, Batzer A, Koehler JA, Polakis P, Schlessinger J, Lydon NB, Moran MF. 1999 Evidence for SH3 domain directed binding and phosphorylation of Sam68 by Src. *Oncogene.* 18, 4647-53
64. Thomas GM, Frame S, Goedert M, Nathke I, Polakis P, Cohen P. 1999 A GSK3- binding peptide from FRAT1 selectively inhibits the GSK3-catalysed phosphorylation of axin and beta-catenin. *FEBS Lett.* 458, 247-51.
65. Peifer M, Polakis P. 2000 Wnt signaling in oncogenesis and embryogenesis—a look outside the nucleus. *Science* 287,1606-9.
66. Polakis P. 2000 Wnt signaling and cancer. *Genes Dev*;14, 1837-1851.
67. Spink KE, Polakis P, Weis WI 2000 Structural basis of the Axin-adenomatous polyposis coli interaction. *EMBO J* 19, 2270-2279.
68. Szeto, W., Jiang, W., Tice, D.A., Rubinfeld, B., Hollingshead, P.G., Fong, S.E., Dugger, D.L., Pham, T., Yansura, D.E., Wong, T.A., Grimaldi, J.C., Corpuz, R.T., Singh J.S., Frantz, G.D., Devaux, B., Crowley, C.W., Schwall, R.H., Eberhard, D.A., Rastelli, L., Polakis, P. and Pennica, D. 2001 Overexpression of the Retinoic Acid-Responsive Gene Stra6 in Human Cancers and its Synergistic Induction by Wnt-1 and Retinoic Acid. *Cancer Res* 61, 4197-4204.
69. Rubinfeld B, Tice DA, Polakis P. 2001 Axin dependent phosphorylation of the adenomatous polyposis coli protein mediated by casein kinase 1 epsilon. *J Biol Chem* 276, 39037-39045.
70. Polakis P. 2001 More than one way to skin a catenin. *Cell* 2001 105, 563-566.
71. Tice DA, Soloviev I, Polakis P. 2002 Activation of the Wnt Pathway Interferes with Serum Response Element-driven Transcription of Immediate Early Genes. *J Biol. Chem.* 277, 6118-6123.
72. Tice DA, Szeto W, Soloviev I, Rubinfeld B, Fong SE, Dugger DL, Winer J,

Williams PM, Wieand D, Smith V, Schwall RH, Pennica D, Polakis P. 2002 Synergistic activation of tumor antigens by wnt-1 signaling and retinoic acid revealed by gene expression profiling. *J Biol Chem.* 277,14329-14335.

73. Polakis, P. 2002 Casein kinase I: A wnt'er of disconnect. *Curr. Biol.* 12, R499.

74. Mao, W., Luis, E., Ross, S., Silva, J., Tan, C., Crowley, C., Chui, C., Franz, G., Senter, P., Koeppen, H., Polakis, P. 2004 EphB2 as a therapeutic antibody drug target for the treatment of colorectal cancer. *Cancer Res.* 64, 781-788.

75. Shibamoto, S., Winer, J., Williams, M., Polakis, P. 2003 A Blockade in Wnt signaling is activated following the differentiation of F9 teratocarcinoma cells. *Exp. Cell Res.* 29211-20.

76. Zhang Y, Eberhard DA, Frantz GD, Dowd P, Wu TD, Zhou Y, Watanabe C, Luoh SM, Polakis P, Hillan KJ, Wood WI, Zhang Z. 2004 GEPIS--quantitative gene expression profiling in normal and cancer tissues. *Bioinformatics*, April 8

## DECLARATION OF PAUL POLAKIS, Ph.D.

I, Paul Polakis, Ph.D., declare and say as follows:

1. I was awarded a Ph.D. by the Department of Biochemistry of the Michigan State University in 1984. My scientific Curriculum Vitae is attached to and forms part of this Declaration (Exhibit A).
2. I am currently employed by Genentech, Inc. where my job title is Staff Scientist. Since joining Genentech in 1999, one of my primary responsibilities has been leading Genentech's Tumor Antigen Project, which is a large research project with a primary focus on identifying tumor cell markers that find use as targets for both the diagnosis and treatment of cancer in humans.
3. As part of the Tumor Antigen Project, my laboratory has been analyzing differential expression of various genes in tumor cells relative to normal cells. The purpose of this research is to identify proteins that are abundantly expressed on certain tumor cells and that are either (i) not expressed, or (ii) expressed at lower levels, on corresponding normal cells. We call such differentially expressed proteins "tumor antigen proteins". When such a tumor antigen protein is identified, one can produce an antibody that recognizes and binds to that protein. Such an antibody finds use in the diagnosis of human cancer and may ultimately serve as an effective therapeutic in the treatment of human cancer.
4. In the course of the research conducted by Genentech's Tumor Antigen Project, we have employed a variety of scientific techniques for detecting and studying differential gene expression in human tumor cells relative to normal cells, at genomic DNA, mRNA and protein levels. An important example of one such technique is the well known and widely used technique of microarray analysis which has proven to be extremely useful for the identification of mRNA molecules that are differentially expressed in one tissue or cell type relative to another. In the course of our research using microarray analysis, we have identified approximately 200 gene transcripts that are present in human tumor cells at significantly higher levels than in corresponding normal human cells. To date, we have generated antibodies that bind to about 30 of the tumor antigen proteins expressed from these differentially expressed gene transcripts and have used these antibodies to quantitatively determine the level of production of these tumor antigen proteins in both human cancer cells and corresponding normal cells. We have then compared the levels of mRNA and protein in both the tumor and normal cells analyzed.
5. From the mRNA and protein expression analyses described in paragraph 4 above, we have observed that there is a strong correlation between changes in the level of mRNA present in any particular cell type and the level of protein

expressed from that mRNA in that cell type. In approximately 80% of our observations we have found that increases in the level of a particular mRNA correlates with changes in the level of protein expressed from that mRNA when human tumor cells are compared with their corresponding normal cells.

6. Based upon my own experience accumulated in more than 20 years of research, including the data discussed in paragraphs 4 and 5 above and my knowledge of the relevant scientific literature, it is my considered scientific opinion that for human genes, an increased level of mRNA in a tumor cell relative to a normal cell typically correlates to a similar increase in abundance of the encoded protein in the tumor cell relative to the normal cell. In fact, it remains a central dogma in molecular biology that increased mRNA levels are predictive of corresponding increased levels of the encoded protein. While there have been published reports of genes for which such a correlation does not exist, it is my opinion that such reports are exceptions to the commonly understood general rule that increased mRNA levels are predictive of corresponding increased levels of the encoded protein.

7. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

Dated: 5/07/04

By: Paul Polakis

Paul Polakis, Ph.D.

## CURRICULUM VITAE

PAUL G. POLAKIS  
Staff Scientist  
Genentech, Inc  
1 DNA Way, MS#40  
S. San Francisco, CA 94080

### EDUCATION:

Ph.D., Biochemistry, Department of Biochemistry,  
Michigan State University (1984)

B.S., Biology. College of Natural Science, Michigan State University (1977)

### PROFESSIONAL EXPERIENCE:

2002-present	Staff Scientist, Genentech, Inc S. San Francisco, CA
1999- 2002	Senior Scientist, Genentech, Inc., S. San Francisco, CA
1997 -1999	Research Director Onyx Pharmaceuticals, Richmond, CA
1992- 1996	Senior Scientist, Project Leader, Onyx Pharmaceuticals, Richmond, CA
1991-1992	Senior Scientist, Chiron Corporation, Emeryville, CA.
1989-1991	Scientist, Cetus Corporation, Emeryville CA.
1987-1989	Postdoctoral Research Associate, Genentech, Inc., South San Francisco, CA.
1985-1987	Postdoctoral Research Associate, Department of Medicine, Duke University Medical Center, Durham, NC

1984-1985

Assistant Professor, Department of Chemistry,  
Oberlin College, Oberlin, Ohio

1980-1984

Graduate Research Assistant, Department of  
Biochemistry, Michigan State University  
East Lansing, Michigan

### **PUBLICATIONS:**

1. Polakis, P. G. and Wilson, J. E. 1982 Purification of a Highly Bindable Rat Brain Hexokinase by High Performance Liquid Chromatography. **Biochem. Biophys. Res. Commun.** 107, 937-943.

2. Polakis, P. G. and Wilson, J. E. 1984 Proteolytic Dissection of Rat Brain Hexokinase: Determination of the Cleavage Pattern during Limited Digestion with Trypsin. **Arch. Biochem. Biophys.** 234, 341-352.

3. Polakis, P. G. and Wilson, J. E. 1985 An Intact Hydrophobic N-Terminal Sequence is Required for the Binding Rat Brain Hexokinase to Mitochondria. **Arch. Biochem. Biophys.** 236, 328-337.

4. Uhing, R. J., Polakis, P. G. and Snyderman, R. 1987 Isolation of GTP-binding Proteins from Myeloid HL60 Cells. **J. Biol. Chem.** 262, 15575-15579.

5. Polakis, P. G., Uhing, R. J. and Snyderman, R. 1988 The Formylpeptide Chemoattractant Receptor Copurifies with a GTP-binding Protein Containing a Distinct 40 kDa Pertussis Toxin Substrate. **J. Biol. Chem.** 263, 4969-4979.

6. Uhing, R. J., Dillon, S., Polakis, P. G., Truett, A. P. and Snyderman, R. 1988 Chemoattractant Receptors and Signal Transduction Processes in Cellular and Molecular Aspects of Inflammation ( Poste, G. and Crooke, S. T. eds.) pp 335-379.

7. Polakis, P. G., Evans, T. and Snyderman 1989 Multiple Chromatographic Forms of the Formylpeptide Chemoattractant Receptor and their Relationship to GTP-binding Proteins. **Biochem. Biophys. Res. Commun.** 161, 276-283.

8. Polakis, P. G., Snyderman, R. and Evans, T. 1989 Characterization of G25K, a GTP-binding Protein Containing a Novel Putative Nucleotide Binding Domain. **Biochem. Biophys. Res. Commun.** 160, 25-32.

9. Polakis, P., Weber, R. F., Nevins, B., Didsbury, J. Evans, T. and Snyderman, R. 1989 Identification of the *ral* and *rac1* Gene Products, Low Molecular Mass GTP-binding Proteins from Human Platelets. **J. Biol. Chem.** 264, 16383-16389.

10. Snyderman, R., Perianin, A., Evans, T., Polakis, P. and Didsbury, J. 1989 G Proteins and Neutrophil Function. In ADP-Ribosylating Toxins and G Proteins: Insights into Signal Transduction. ( J. Moss and M. Vaughn, eds.) Amer. Soc. Microbiol. pp. 295-323.

11. Hart, M.J., Polakis, P.G., Evans, T. and Cerrione, R.A. 1990 The Identification and Characterization of an Epidermal Growth Factor-Stimulated Phosphorylation of a Specific Low Molecular Mass GTP-binding Protein in a Reconstituted Phospholipid Vesicle System. **J. Biol. Chem.** 265, 5990-6001.
12. Yatani, A., Okabe, K., Polakis, P., Halenbeck, R., McCormick, F. and Brown, A. M. 1990 ras p21 and GAP Inhibit Coupling of Muscarinic Receptors to Atrial K<sup>+</sup> Channels. **Cell**. 61, 769-776.
13. Munemitsu, S., Innis, M.A., Clark, R., McCormick, F., Ullrich, A. and Polakis, P.G. 1990 Molecular Cloning and Expression of a G25K cDNA, the Human Homolog of the Yeast Cell Cycle Gene CDC42. **Mol. Cell. Biol.** 10, 5977-5982.
14. Polakis, P.G., Rubinfeld, B., Evans, T. and McCormick, F. 1991 Purification of Plasma Membrane-Associated GTPase Activating Protein (GAP) Specific for rap-1/krev-1 from HL60 Cells. **Proc. Natl. Acad. Sci. USA** 88, 239-243.
15. Moran, M. F., Polakis, P., McCormick, F., Pawson, T. and Ellis, C. 1991 Protein Tyrosine Kinases Regulate the Phosphorylation, Protein Interactions, Subcellular Distribution, and Activity of p21ras GTPase Activating Protein. **Mol. Cell. Biol.** 11, 1804-1812.
16. Rubinfeld, B., Wong, G., Bekesi, E., Wood, A., McCormick, F. and Polakis, P. G. 1991 A Synthetic Peptide Corresponding to a Sequence in the GTPase Activating Protein Inhibits p21<sup>ras</sup> Stimulation and Promotes Guanine Nucleotide Exchange. **Internatl. J. Peptide and Prot. Res.** 38, 47-53.
17. Rubinfeld, B., Munemitsu, S., Clark, R., Conroy, L., Watt, K., Crosier, W., McCormick, F., and Polakis, P. 1991 Molecular Cloning of a GTPase Activating Protein Specific for the Krev-1 Protein p21<sup>rap1</sup>. **Cell** 65, 1033-1042.
18. Zhang, K., Papageorge, A., G., Martin, P., Vass, W. C., Olah, Z., Polakis, P., McCormick, F. and Lowy, D. R. 1991 Heterogenous Amino Acids in RAS and Rap1A Specifying Sensitivity to GAP Proteins. **Science** 254, 1630-1634.
19. Martin, G., Yatani, A., Clark, R., Polakis, P., Brown, A. M. and McCormick, F. 1992 GAP Domains Responsible for p21<sup>ras</sup>-dependent Inhibition of Muscarinic Atrial K<sup>+</sup> Channel Currents. **Science** 255, 192-194.
20. McCormick, F., Martin, G. A., Clark, R., Bollag, G. and Polakis, P. 1992 Regulation of p21ras by GTPase Activating Proteins. Cold Spring Harbor **Symposia on Quantitative Biology**. Vol. 56, 237-241.
21. Pronk, G. B., Polakis, P., Wong, G., deVries-Smits, A. M., Bos J. L. and McCormick, F. 1992 p60<sup>v-src</sup> Can Associate with and Phosphorylate the p21<sup>ras</sup> GTPase Activating Protein. **Oncogene** 7, 389-394.
22. Polakis P. and McCormick, F. 1992 Interactions Between p21<sup>ras</sup> Proteins and Their GTPase Activating Proteins. In **Cancer Surveys** ( Franks, L. M., ed.) 12, 25-42.



23. Wong, G., Muller, O., Clark, R., Conroy, L., Moran, M., Polakis, P. and McCormick, F. 1992 Molecular cloning and nucleic acid binding properties of the GAP-associated tyrosine phosphoprotein p62. *Cell* 69, 551-558.
24. Polakis, P., Rubinfeld, B. and McCormick, F. 1992 Phosphorylation of rap1GAP in vivo and by cAMP-dependent Kinase and the Cell Cycle p34<sup>cdc2</sup> Kinase in vitro. *J. Biol. Chem.* 267, 10780-10785.
25. McCabe, P.C., Haubrauck, H., Polakis, P., McCormick, F., and Innis, M. A. 1992 Functional Interactions Between p21<sup>rap1A</sup> and Components of the Budding pathway of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 12, 4084-4092.
26. Rubinfeld, B., Crosier, W.J., Albert, I., Conroy, L., Clark, R., McCormick, F. and Polakis, P. 1992 Localization of the rap1GAP Catalytic Domain and Sites of Phosphorylation by Mutational Analysis. *Mol. Cell. Biol.* 12, 4634-4642.
27. Ando, S., Kaibuchi, K., Sasaki, K., Hiraoka, T., Nishiyama, T., Mizuno, T., Asada, M., Nunoi, H., Matsuda, I., Matsuura, Y., Polakis, P., McCormick, F. and Takai, Y. 1992 Post-translational processing of rac p21s is important both for their interaction with the GDP/GTP exchange proteins and for their activation of NADPH oxidase. *J. Biol. Chem.* 267, 25709-25713.
28. Janoueix-Lerosey, I., Polakis, P., Tavitian, A. and deGunzberg, J. 1992 Regulation of the GTPase activity of the ras-related rap2 protein. *Biochem. Biophys. Res. Commun.* 189, 455-464.
29. Polakis, P. 1993 GAPs Specific for the rap1/Krev-1 Protein. in GTP-binding Proteins: the ras-superfamily. ( J.C. LaCale and F. McCormick, eds.) 445-452.
30. Polakis, P. and McCormick, F. 1993 Structural requirements for the interaction of p21<sup>ras</sup> with GAP, exchange factors, and its biological effector target. *J. Biol. Chem.* 268, 9157-9160.
31. Rubinfeld, B., Souza, B., Albert, I., Muller, O., Chamberlain, S., Masiarz, F., Munemitsu, S. and Polakis, P. 1993 Association of the APC gene product with beta- catenin. *Science* 262, 1731-1734.
32. Weiss, J., Rubinfeld, B., Polakis, P., McCormick, F., Cavenee, W. A. and Arden, K. 1993. The gene for human rap1-GTPase activating protein (rap1GAP) maps to chromosome 1p35-1p36.1. *Cytogenet. Cell Genet.* 66, 18-21.
33. Sato, K. Y., Polakis, P., Haubruck, H., Fasching, C. L., McCormick, F. and Stanbridge, E. J. 1994 Analysis of the tumor suppressor activity of the K-rev gene in human tumor cell lines. *Cancer Res.* 54, 552-559.
34. Janoueix-Lerosey, I., Fontenay, M., Tobelem, G., Tavitian, A., Polakis, P. and DeGunzburg, J. 1994 Phosphorylation of rap1GAP during the cell cycle. *Biochem. Biophys. Res. Commun.* 202, 967-975
35. Munemitsu, S., Souza, B., Mueller, O., Albert, I., Rubinfeld, B., and Polakis, P. 1994 The APC gene product associates with microtubules in vivo and affects their assembly in vitro. *Cancer Res.* 54, 3676-3681.

36. Rubinfeld, B. and Polakis, P. 1995 Purification of baculovirus produced rap1GAP. **Methods Enz.** 255,31
37. Polakis, P. 1995 Mutations in the APC gene and their implications for protein structure and function. **Current Opinions in Genetics and Development** 5, 66-71
38. Rubinfeld, B., Souza, B., Albert, I., Munemitsu, S. and Polakis P. 1995 The APC protein and E-cadherin form similar but independent complexes with  $\alpha$ -catenin,  $\beta$ -catenin and Plakoglobin. **J. Biol. Chem.** 270, 5549-5555
39. Munemitsu, S., Albert, I., Souza, B., Rubinfeld, B., and Polakis, P. 1995 Regulation of intracellular  $\beta$ -catenin levels by the APC tumor suppressor gene. **Proc. Natl. Acad. Sci.** 92, 3046-3050.
40. Lock, P., Fumagalli, S., Polakis, P. McCormick, F. and Courtneidge, S. A. 1996 The human p62 cDNA encodes Sam68 and not the rasGAP-associated p62 protein. **Cell** 84, 23-24.
41. Papkoff, J., Rubinfeld, B., Schryver, B. and Polakis, P. 1996 Wnt-1 regulates free pools of catenins and stabilizes APC-catenin complexes. **Mol. Cell. Biol.** 16, 2128-2134.
42. Rubinfeld, B., Albert, I., Porfiri, E., Fiol, C., Munemitsu, S. and Polakis, P. 1996 Binding of GSK3 $\beta$  to the APC- $\beta$ -catenin complex and regulation of complex assembly. **Science** 272, 1023-1026.
43. Munemitsu, S., Albert, I., Rubinfeld, B. and Polakis, P. 1996 Deletion of amino-terminal structure stabilizes  $\beta$ -catenin in vivo and promotes the hyperphosphorylation of the APC tumor suppressor protein. **Mol. Cell. Biol.** 16, 4088-4094.
44. Hart, M. J., Callow, M. G., Sousa, B. and Polakis P. 1996 IQGAP1, a calmodulin binding protein with a rasGAP related domain, is a potential effector for cdc42Hs. **EMBO J.** 15, 2997-3005.
45. Nathke, I. S., Adams, C. L., Polakis, P., Sellin, J. and Nelson, W. J. 1996 The adenomatous polyposis coli (APC) tumor suppressor protein is localized to plasma membrane sites involved in active epithelial cell migration. **J. Cell. Biol.** 134, 165-180.
46. Hart, M. J., Sharma, S., elMasry, N., Qui, R-G., McCabe, P., Polakis, P. and Bollag, G. 1996 Identification of a novel guanine nucleotide exchange factor for the rho GTPase. **J. Biol. Chem.** 271, 25452.
47. Thomas JE, Smith M, Rubinfeld B, Gutowski M, Beckmann RP, and Polakis P. 1996 Subcellular localization and analysis of apparent 180-kDa and 220-kDa proteins of the breast cancer susceptibility gene, BRCA1. **J. Biol. Chem.** 1996 271, 28630-28635
48. Hayashi, S., Rubinfeld, B., Souza, B., Polakis, P., Wieschaus, E., and Levine, A. 1997 A Drosophila homolog of the tumor suppressor adenomatous polyposis coli

down-regulates  $\beta$ -catenin but its zygotic expression is not essential for the regulation of armadillo. **Proc. Natl. Acad. Sci.** 94, 242-247.

49. Vleminckx, K., Rubinfeld, B., Polakis, P. and Gumbiner, B. 1997 The APC tumor suppressor protein induces a new axis in *Xenopus* embryos. **J. Cell. Biol.** 136, 411-420.

50. Rubinfeld, B., Robbins, P., El-Gamil, M., Albert, I., Porfiri, P. and Polakis, P. 1997 Stabilization of  $\beta$ -catenin by genetic defects in melanoma cell lines. **Science** 275, 1790-1792.

51. Polakis, P. The adenomatous polyposis coli (APC) tumor suppressor. 1997 **Biochem. Biophys. Acta**, 1332, F127-F147.

52. Rubinfeld, B., Albert, I., Porfiri, E., Munemitsu, S., and Polakis, P. 1997 Loss of  $\beta$ -catenin regulation by the APC tumor suppressor protein correlates with loss of structure due to common somatic mutations of the gene. **Cancer Res.** 57, 4624-4630.

53. Porfiri, E., Rubinfeld, B., Albert, I., Hovanes, K., Waterman, M., and Polakis, P. 1997 Induction of a  $\beta$ -catenin-LEF-1 complex by wnt-1 and transforming mutants of  $\beta$ -catenin. **Oncogene** 15, 2833-2839.

54. Thomas JE, Smith M, Tonkinson JL, Rubinfeld B, and Polakis P., 1997 Induction of phosphorylation on BRCA1 during the cell cycle and after DNA damage. **Cell Growth Differ.** 8, 801-809.

55. Hart, M., de los Santos, R., Albert, I., Rubinfeld, B., and Polakis P., 1998 Down regulation of  $\beta$ -catenin by human Axin and its association with the adenomatous polyposis coli (APC) tumor suppressor,  $\beta$ -catenin and glycogen synthase kinase  $\beta$ . **Current Biology** 8, 573-581.

56. Polakis, P. 1998 The oncogenic activation of  $\beta$ -catenin. **Current Opinions in Genetics and Development** 9, 15-21

57. Matt Hart, Jean-Paul Concordet, Irina Lassot, Iris Albert, Rico del los Santos, Herve Durand, Christine Perret, Bonnee Rubinfeld, Florence Margottin, Richard Benarous and Paul Polakis. 1999 The F-box protein  $\beta$ -TrCP associates with phosphorylated  $\beta$ -catenin and regulates its activity in the cell. **Current Biology** 9, 207-10.

58. Howard C. Crawford, Barbara M. Fingleton, Bonnee Rubinfeld, Paul Polakis and Lynn M. Matrisian 1999 The metalloproteinase matrilysin is a target of  $\beta$ -catenin transactivation in intestinal tumours. **Oncogene** 18, 2883-91.

59. Meng J, Glick JL, Polakis P, Casey PJ. 1999 Functional interaction between G $\alpha$ (z) and Rap1GAP suggests a novel form of cellular cross-talk. **J Biol Chem.** 17, 36663-9

60. Vijayasurian Easwaran, Virginia Song, **Paul Polakis** and Steve Myers 1999 The ubiquitin-proteasome pathway and serine kinase activity modulate APC mediated regulation of  $\beta$ -catenin-LEF signaling. **J. Biol. Chem.** 274(23):16641-5.
61. **Polakis P**, Hart M and Rubinfeld B. 1999 Defects in the regulation of beta-catenin in colorectal cancer. **Adv Exp Med Biol.** 470, 23-32
62. Shen Z, Batzer A, Koehler JA, **Polakis P**, Schlessinger J, Lydon NB, Moran MF. 1999 Evidence for SH3 domain directed binding and phosphorylation of Sam68 by Src. **Oncogene.** 18, 4647-53
64. Thomas GM, Frame S, Goedert M, Nathke I, **Polakis P**, Cohen P. 1999 A GSK3- binding peptide from FRAT1 selectively inhibits the GSK3-catalysed phosphorylation of axin and beta-catenin. **FEBS Lett.** 458, 247-51.
65. Peifer M, **Polakis P**. 2000 Wnt signaling in oncogenesis and embryogenesis--a look outside the nucleus. **Science** 287,1606-9.
66. **Polakis P**. 2000 Wnt signaling and cancer. **Genes Dev**;14, 1837-1851.
67. Spink KE, **Polakis P**, Weis WI. 2000 Structural basis of the Axin-adenomatous polyposis coli interaction. **EMBO J** 19, 2270-2279.
68. Szeto, W., Jiang, W., Tice, D.A., Rubinfeld, B., Hollingshead, P.G., Fong, S.E., Dugger, D.L., Pham, T., Yansura, D.E., Wong, T.A., Grimaldi, J.C., Corpuz, R.T., Singh J.S., Frantz, G.D., Devaux, B., Crowley, C.W., Schwall, R.H., Eberhard, D.A., Rastelli, L., **Polakis, P.** and Pennica, D. 2001 Overexpression of the Retinoic Acid-Responsive Gene Stra6 in Human Cancers and its Synergistic Induction by Wnt-1 and Retinoic Acid. **Cancer Res** 61, 4197-4204.
69. Rubinfeld B, Tice DA, **Polakis P**. 2001 Axin dependent phosphorylation of the adenomatous polyposis coli protein mediated by casein kinase 1 epsilon. **J Biol Chem** 276, 39037-39045.
70. **Polakis P**. 2001 More than one way to skin a catenin. **Cell** 2001 105, 563-566.
71. Tice DA, Soloviev I, **Polakis P**. 2002 Activation of the Wnt Pathway Interferes with Serum Response Element-driven Transcription of Immediate Early Genes. **J Biol. Chem.** 277, 6118-6123.
72. Tice DA, Szeto W, Soloviev I, Rubinfeld B, Fong SE, Dugger DL, Winer J,

- Williams PM, Wieand D, Smith v, Schwall RH, Pennica D, Polakis P. 2002 Synergistic activation of tumor antigens by wnt-1 signaling and retinoic acid revealed by gene expression profiling. **J Biol Chem.** 277,14329-14335.
73. Polakis, P. 2002 Casein kinase I: A wnt'er of disconnect. **Curr. Biol.** 12, R499.
74. Mao, W., Luis, E., Ross, S., Silva, J., Tan, C., Crowley, C., Chui, C., Franz, G., Senter, P., Koeppen, H., Polakis, P. 2004 EphB2 as a therapeutic antibody drug target for the treatment of colorectal cancer. **Cancer Res.** 64, 781-788.
75. Shibamoto, S., Winer, J., Williams, M., Polakis, P. 2003 A Blockade in Wnt signaling is activated following the differentiation of F9 teratocarcinoma cells. **Exp. Cell Res.** 29211-20.
76. Zhang Y, Eberhard DA, Frantz GD, Dowd P, Wu TD, Zhou Y, Watanabe C, Luoh SM, Polakis P, Hillan KJ, Wood WI, Zhang Z. 2004 GEPIS—quantitative gene expression profiling in normal and cancer tissues. **Bioinformatics**, April 8

SECOND DECLARATION OF PAUL POLAKIS, Ph.D.

I, Paul Polakis, Ph.D., declare and say as follows:

1. I am currently employed by Genentech, Inc. where my job title is Staff Scientist.
2. Since joining Genentech in 1999, one of my primary responsibilities has been leading Genentech's Tumor Antigen Project, which is a large research project with a primary focus on identifying tumor cell markers that find use as targets for both the diagnosis and treatment of cancer in humans.
3. As I stated in my previous Declaration dated May 7, 2004 (attached as Exhibit A), my laboratory has been employing a variety of techniques, including microarray analysis, to identify genes which are differentially expressed in human tumor tissue relative to normal human tissue. The primary purpose of this research is to identify proteins that are abundantly expressed on certain human tumor tissue(s) and that are either (i) not expressed, or (ii) expressed at detectably lower levels, on normal tissue(s).
4. In the course of our research using microarray analysis, we have identified approximately 200 gene transcripts that are present in human tumor tissue at significantly higher levels than in normal human tissue. To date, we have successfully generated antibodies that bind to 31 of the tumor antigen proteins expressed from these differentially expressed gene transcripts and have used these antibodies to quantitatively determine the level of production of these tumor antigen proteins in both human tumor tissue and normal tissue. We have then quantitatively compared the levels of mRNA and protein in both the tumor and normal tissues analyzed. The results of these analyses are attached herewith as Exhibit B. In Exhibit B, "+" means that the mRNA or protein was detectably overexpressed in the tumor tissue relative to normal tissue and "-" means that no detectable overexpression was observed in the tumor tissue relative to normal tissue.
5. As shown in Exhibit B, of the 31 genes identified as being detectably overexpressed in human tumor tissue as compared to normal human tissue at the mRNA level, 28 of them (i.e., greater than 90%) are also detectably overexpressed in human tumor tissue as compared to normal human tissue at the protein level. As such, in the cases where we have been able to quantitatively measure both (i) mRNA and (ii) protein levels in both (i) tumor tissue and (ii) normal tissue, we have observed that in the vast majority of cases, there is a very strong correlation between increases in mRNA expression and increases in the level of protein encoded by that mRNA.

6. Based upon my own experience accumulated in more than 20 years of research, including the data discussed in paragraphs 4-5 above and my knowledge of the relevant scientific literature, it is my considered scientific opinion that for human genes, an increased level of mRNA in a tumor tissue relative to a normal tissue more often than not correlates to a similar increase in abundance of the encoded protein in the tumor tissue relative to the normal tissue. In fact, it remains a generally accepted working assumption in molecular biology that increased mRNA levels are more often than not predictive of elevated levels of the encoded protein. In fact, an entire industry focusing on the research and development of therapeutic antibodies to treat a variety of human diseases, such as cancer, operates on this working assumption.
7. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

Dated: 3-29-00

By: Paul Polakis

Paul Polakis, Ph.D.

DECLARATION OF PAUL POLAKIS, Ph.D.

I, Paul Polakis, Ph.D., declare and say as follows:

1. I was awarded a Ph.D. by the Department of Biochemistry of the Michigan State University in 1984. My scientific Curriculum Vitae is attached to and forms part of this Declaration (Exhibit A).
2. I am currently employed by Genentech, Inc. where my job title is Staff Scientist. Since joining Genentech in 1999, one of my primary responsibilities has been leading Genentech's Tumor Antigen Project, which is a large research project with a primary focus on identifying tumor cell markers that find use as targets for both the diagnosis and treatment of cancer in humans.
3. As part of the Tumor Antigen Project, my laboratory has been analyzing differential expression of various genes in tumor cells relative to normal cells. The purpose of this research is to identify proteins that are abundantly expressed on certain tumor cells and that are either (i) not expressed, or (ii) expressed at lower levels, on corresponding normal cells. We call such differentially expressed proteins "tumor antigen proteins". When such a tumor antigen protein is identified, one can produce an antibody that recognizes and binds to that protein. Such an antibody finds use in the diagnosis of human cancer and may ultimately serve as an effective therapeutic in the treatment of human cancer.
4. In the course of the research conducted by Genentech's Tumor Antigen Project, we have employed a variety of scientific techniques for detecting and studying differential gene expression in human tumor cells relative to normal cells, at genomic DNA, mRNA and protein levels. An important example of one such technique is the well known and widely used technique of microarray analysis which has proven to be extremely useful for the identification of mRNA molecules that are differentially expressed in one tissue or cell type relative to another. In the course of our research using microarray analysis, we have identified approximately 200 gene transcripts that are present in human tumor cells at significantly higher levels than in corresponding normal human cells. To date, we have generated antibodies that bind to about 30 of the tumor antigen proteins expressed from these differentially expressed gene transcripts and have used these antibodies to quantitatively determine the level of production of these tumor antigen proteins in both human cancer cells and corresponding normal cells. We have then compared the levels of mRNA and protein in both the tumor and normal cells analyzed.
5. From the mRNA and protein expression analyses described in paragraph 4 above, we have observed that there is a strong correlation between changes in the level of mRNA present in any particular cell type and the level of protein



expressed from that mRNA in that cell type. In approximately 80% of our observations we have found that increases in the level of a particular mRNA correlates with changes in the level of protein expressed from that mRNA when human tumor cells are compared with their corresponding normal cells.

6. Based upon my own experience accumulated in more than 20 years of research, including the data discussed in paragraphs 4 and 5 above and my knowledge of the relevant scientific literature, it is my considered scientific opinion that for human genes, an increased level of mRNA in a tumor cell relative to a normal cell typically correlates to a similar increase in abundance of the encoded protein in the tumor cell relative to the normal cell. In fact, it remains a central dogma in molecular biology that increased mRNA levels are predictive of corresponding increased levels of the encoded protein. While there have been published reports of genes for which such a correlation does not exist, it is my opinion that such reports are exceptions to the commonly understood general rule that increased mRNA levels are predictive of corresponding increased levels of the encoded protein.

7. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

Dated: 5/07/04

By: Paul Polakis

Paul Polakis, Ph.D.

## CURRICULUM VITAE

PAUL G. POLAKIS  
Staff Scientist  
Genentech, Inc  
1 DNA Way, MS#40  
S. San Francisco, CA 94080

### EDUCATION:

Ph.D., Biochemistry, Department of Biochemistry,  
Michigan State University (1984)

B.S., Biology. College of Natural Science, Michigan State University (1977)

### PROFESSIONAL EXPERIENCE:

2002-present	Staff Scientist, Genentech, Inc S. San Francisco, CA
1999- 2002	Senior Scientist, Genentech, Inc., S. San Francisco, CA
1997 -1999	Research Director Onyx Pharmaceuticals, Richmond, CA
1992- 1996	Senior Scientist, Project Leader, Onyx Pharmaceuticals, Richmond, CA
1991-1992	Senior Scientist, Chiron Corporation, Emeryville, CA.
1989-1991	Scientist, Cetus Corporation, Emeryville CA.
1987-1989	Postdoctoral Research Associate, Genentech, Inc., South San Francisco, CA.
1985-1987	Postdoctoral Research Associate, Department of Medicine, Duke University Medical Center, Durham, NC

1984-1985

Assistant Professor, Department of Chemistry,  
Oberlin College, Oberlin, Ohio

1980-1984

Graduate Research Assistant, Department of  
Biochemistry, Michigan State University  
East Lansing, Michigan

**PUBLICATIONS:**

1. Polakis, P. G. and Wilson, J. E. 1982 Purification of a Highly Bindable Rat Brain Hexokinase by High Performance Liquid Chromatography. *Biochem. Biophys. Res. Commun.* 107, 937-943.
2. Polakis, P.G. and Wilson, J. E. 1984 Proteolytic Dissection of Rat Brain Hexokinase: Determination of the Cleavage Pattern during Limited Digestion with Trypsin. *Arch. Biochem. Biophys.* 234, 341-352.
3. Polakis, P. G. and Wilson, J. E. 1985 An Intact Hydrophobic N-Terminal Sequence is Required for the Binding Rat Brain Hexokinase to Mitochondria. *Arch. Biochem. Biophys.* 236, 328-337.
4. Uhing, R.J., Polakis, P.G. and Snyderman, R. 1987 Isolation of GTP-binding Proteins from Myeloid HL60 Cells. *J. Biol. Chem.* 262, 15575-15579.
5. Polakis, P.G., Uhing, R.J. and Snyderman, R. 1988 The Formylpeptide Chemoattractant Receptor Copurifies with a GTP-binding Protein Containing a Distinct 40 kDa Pertussis Toxin Substrate. *J. Biol. Chem.* 263, 4969-4979.
6. Uhing, R. J., Dillon, S., Polakis, P. G., Truett, A. P. and Snyderman, R. 1988 Chemoattractant Receptors and Signal Transduction Processes in Cellular and Molecular Aspects of Inflammation ( Poste, G. and Crooke, S. T. eds.) pp 335-379.
7. Polakis, P.G., Evans, T. and Snyderman 1989 Multiple Chromatographic Forms of the Formylpeptide Chemoattractant Receptor and their Relationship to GTP-binding Proteins. *Biochem. Biophys. Res. Commun.* 161, 276-283.
8. Polakis, P. G., Snyderman, R. and Evans, T. 1989 Characterization of G25K, a GTP-binding Protein Containing a Novel Putative Nucleotide Binding Domain. *Biochem. Biophys. Res. Commun.* 160, 25-32.
9. Polakis, P., Weber, R.F., Nevins, B., Didsbury, J. Evans, T. and Snyderman, R. 1989 Identification of the *ral* and *rac1* Gene Products, Low Molecular Mass GTP-binding Proteins from Human Platelets. *J. Biol. Chem.* 264, 16383-16389.
10. Snyderman, R., Perianin, A., Evans, T., Polakis, P. and Didsbury, J. 1989 G Proteins and Neutrophil Function. In *ADP-Ribosylating Toxins and G Proteins: Insights into Signal Transduction.* ( J. Moss and M. Vaughn, eds.) Amer. Soc. Microbiol. pp. 295-323.

11. Hart, M.J., Polakis, P.G., Evans, T. and Cerrione, R.A. 1990 The Identification and Characterization of an Epidermal Growth Factor-Stimulated Phosphorylation of a Specific Low Molecular Mass GTP-binding Protein in a Reconstituted Phospholipid Vesicle System. *J. Biol. Chem.* 265, 5990-6001.
12. Yatani, A., Okabe, K., Polakis, P., Halenbeck, R., McCormick, F. and Brown, A. M. 1990 ras p21 and GAP Inhibit Coupling of Muscarinic Receptors to Atrial K<sup>+</sup> Channels. *Cell*. 61, 769-776.
13. Munemitsu, S., Innis, M.A., Clark, R., McCormick, F., Ullrich, A. and Polakis, P.G. 1990 Molecular Cloning and Expression of a G25K cDNA, the Human Homolog of the Yeast Cell Cycle Gene CDC42. *Mol. Cell. Biol.* 10, 5977-5982.
14. Polakis, P.G., Rubinfeld, B., Evans, T. and McCormick, F. 1991 Purification of Plasma Membrane-Associated GTPase Activating Protein (GAP) Specific for rap-1/krev-1 from HL60 Cells. *Proc. Natl. Acad. Sci. USA* 88, 239-243.
15. Moran, M. F., Polakis, P., McCormick, F., Pawson, T. and Ellis, C. 1991 Protein Tyrosine Kinases Regulate the Phosphorylation, Protein Interactions, Subcellular Distribution, and Activity of p21ras GTPase Activating Protein. *Mol. Cell. Biol.* 11, 1804-1812.
16. Rubinfeld, B., Wong, G., Bekesi, E., Wood, A., McCormick, F. and Polakis, P. G. 1991 A Synthetic Peptide Corresponding to a Sequence in the GTPase Activating Protein Inhibits p21<sup>ras</sup> Stimulation and Promotes Guanine Nucleotide Exchange. *Internatl. J. Peptide and Prot. Res.* 38, 47-53.
17. Rubinfeld, B., Munemitsu, S., Clark, R., Conroy, L., Watt, K., Crosier, W., McCormick, F., and Polakis, P. 1991 Molecular Cloning of a GTPase Activating Protein Specific for the Krev-1 Protein p21<sup>rap1</sup>. *Cell* 65, 1033-1042.
18. Zhang, K., Papageorge, A., G., Martin, P., Vass, W. C., Olah, Z., Polakis, P., McCormick, F. and Lowy, D. R. 1991 Heterogenous Amino Acids in RAS and Rap1A Specifying Sensitivity to GAP Proteins. *Science* 254, 1630-1634.
19. Martin, G., Yatani, A., Clark, R., Polakis, P., Brown, A. M. and McCormick, F. 1992 GAP Domains Responsible for p21<sup>ras</sup>-dependent Inhibition of Muscarinic Atrial K<sup>+</sup> Channel Currents. *Science* 255, 192-194.
20. McCormick, F., Martin, G. A., Clark, R., Bollag, G. and Polakis, P. 1992 Regulation of p21ras by GTPase Activating Proteins. *Cold Spring Harbor Symposia on Quantitative Biology*. Vol. 56, 237-241.
21. Pronk, G. B., Polakis, P., Wong, G., deVries-Smits, A. M., Bos J. L. and McCormick, F. 1992 p60<sup>v-src</sup> Can Associate with and Phosphorylate the p21<sup>ras</sup> GTPase Activating Protein. *Oncogene* 7, 389-394.
22. Polakis P. and McCormick, F. 1992 Interactions Between p21<sup>ras</sup> Proteins and Their GTPase Activating Proteins. In Cancer Surveys ( Franks, L. M., ed.) 12, 25-42.

23. Wong, G., Muller, O., Clark, R., Conroy, L., Moran, M., Polakis, P. and McCormick, F. 1992 Molecular cloning and nucleic acid binding properties of the GAP-associated tyrosine phosphoprotein p62. *Cell* 69, 551-558.
24. Polakis, P., Rubinfeld, B. and McCormick, F. 1992 Phosphorylation of rap1GAP in vivo and by cAMP-dependent Kinase and the Cell Cycle p34<sup>cdc2</sup> Kinase in vitro. *J. Biol. Chem.* 267, 10780-10785.
25. McCabe, P.C., Haubrauck, H., Polakis, P., McCormick, F., and Innis, M. A. 1992 Functional Interactions Between p21<sup>rap1A</sup> and Components of the Budding pathway of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 12, 4084-4092.
26. Rubinfeld, B., Crosier, W.J., Albert, I., Conroy, L., Clark, R., McCormick, F. and Polakis, P. 1992 Localization of the rap1GAP Catalytic Domain and Sites of Phosphorylation by Mutational Analysis. *Mol. Cell. Biol.* 12, 4634-4642.
27. Ando, S., Kaibuchi, K., Sasaki, K., Hiraoka, T., Nishiyama, T., Mizuno, T., Asada, M., Nunoi, H., Matsuda, I., Matsuura, Y., Polakis, P., McCormick, F. and Takai, Y. 1992 Post-translational processing of rac p21s is important both for their interaction with the GDP/GTP exchange proteins and for their activation of NADPH oxidase. *J. Biol. Chem.* 267, 25709-25713.
28. Janoueix-Lerosey, I., Polakis, P., Tavitian, A. and deGünzberg, J. 1992 Regulation of the GTPase activity of the ras-related rap2 protein. *Biochem. Biophys. Res. Commun.* 189, 455-464.
29. Polakis, P. 1993 GAPs Specific for the rap1/Krev-1 Protein. in GTP-binding Proteins: the ras-superfamily. ( J.C. LaCale and F. McCormick, eds.) 445-452.
30. Polakis, P. and McCormick, F. 1993 Structural requirements for the interaction of p21<sup>ras</sup> with GAP, exchange factors, and its biological effector target. *J. Biol. Chem.* 268, 9157-9160.
31. Rubinfeld, B., Souza, B., Albert, I., Muller, O., Chamberlain, S., Masiarz, F., Munemitsu, S. and Polakis, P. 1993 Association of the APC gene product with beta-catenin. *Science* 262, 1731-1734.
32. Weiss, J., Rubinfeld, B., Polakis, P., McCormick, F., Cavenee, W. A. and Arden, K. 1993 The gene for human rap1-GTPase activating protein (rap1GAP) maps to chromosome 1p35-1p36.1. *Cytogenet. Cell-Genet.* 66, 18-21.
- 
33. Sato, K. Y., Polakis, P., Haubruck, H., Fasching, C. L., McCormick, F. and Stanbridge, E. J. 1994 Analysis of the tumor suppressor activity of the K-rev gene in human tumor cell lines. *Cancer Res.* 54, 552-559.
34. Janoueix-Lerosey, I., Fontenay, M., Tobelem, G., Tavitian, A., Polakis, P. and DeGünzburg, J. 1994 Phosphorylation of rap1GAP during the cell cycle. *Biochem. Biophys. Res. Commun.* 202, 967-975.
35. Munemitsu, S., Souza, B., Mueller, O., Albert, I., Rubinfeld, B., and Polakis, P. 1994 The APC gene product associates with microtubules in vivo and affects their assembly in vitro. *Cancer Res.* 54, 3676-3681.

36. Rubinfeld, B. and Polakis, P. 1995 Purification of baculovirus produced rap1GAP. *Methods Enz.* 255,31
37. Polakis, P. 1995 Mutations in the APC gene and their implications for protein structure and function: *Current Opinions in Genetics and Development* 5, 66-71
38. Rubinfeld, B., Souza, B., Albert, I., Munemitsu, S. and Polakis P. 1995 The APC protein and E-cadherin form similar but independent complexes with  $\alpha$ -catenin,  $\beta$ -catenin and Plakoglobin. *J. Biol. Chem.* 270, 5549-5555
39. Munemitsu, S., Albert, I., Souza, B., Rubinfeld, B., and Polakis, P. 1995 Regulation of intracellular  $\beta$ -catenin levels by the APC tumor suppressor gene. *Proc. Natl. Acad. Sci.* 92, 3046-3050.
40. Lock, P., Fumagalli, S., Polakis, P. McCormick, F. and Courtneidge, S. A. 1996 The human p62 cDNA encodes Sam68 and not the rasGAP-associated p62 protein. *Cell* 84, 23-24.
41. Papkoff, J., Rubinfeld, B., Schryver, B. and Polakis, P. 1996 Wnt-1 regulates free pools of catenins and stabilizes APC-catenin complexes. *Mol. Cell. Biol.* 16, 2128-2134.
42. Rubinfeld, B., Albert, I., Porfiri, E., Fiol, C., Munemitsu, S. and Polakis, P. 1996 Binding of GSK3 $\beta$  to the APC- $\beta$ -catenin complex and regulation of complex assembly. *Science* 272, 1023-1026.
43. Munemitsu, S., Albert, I., Rubinfeld, B. and Polakis, P. 1996 Deletion of amino-terminal structure stabilizes  $\beta$ -catenin in vivo and promotes the hyperphosphorylation of the APC tumor suppressor protein. *Mol. Cell. Biol.* 16, 4088-4094.
44. Hart, M. J., Callow, M. G., Sousa, B. and Polakis P. 1996 IQGAP1, a calmodulin binding protein with a rasGAP related domain, is a potential effector for cdc42Hs. *EMBO J.* 15, 2997-3005.
45. Nathke, I. S., Adams, C. L., Polakis, P., Sellin, J. and Nelson, W. J. 1996 The adenomatous polyposis coli (APC) tumor suppressor protein is localized to plasma membrane sites involved in active epithelial cell migration. *J. Cell. Biol.* 134, 165-180.
- 
46. Hart, M. J., Sharma, S., elMasry, N., Qui, R-G., McCabe, P., Polakis, P. and Bollag, G. 1996 Identification of a novel guanine nucleotide exchange factor for the rho GTPase. *J. Biol. Chem.* 271, 25452.
47. Thomas JE, Smith M, Rubinfeld B, Gutowski M, Beckmann RP, and Polakis P. 1996 Subcellular localization and analysis of apparent 180-kDa and 220-kDa proteins of the breast cancer susceptibility gene, BRCA1. *J. Biol. Chem.* 1996 271, 28630-28635
48. Hayashi, S., Rubinfeld, B., Souza, B., Polakis, P., Wieschaus, E., and Levine, A. 1997 A Drosophila homolog of the tumor suppressor adenomatous polyposis coli

down-regulates  $\beta$ -catenin but its zygotic expression is not essential for the regulation of armadillo. *Proc. Natl. Acad. Sci.* 94, 242-247.

49. Vleminckx, K., Rubinfeld, B., Polakis, P. and Gumbiner, B. 1997 The APC tumor suppressor protein induces a new axis in *Xenopus* embryos. *J. Cell. Biol.* 136, 411-420.

50. Rubinfeld, B., Robbins, P., El-Gamil, M., Albert, I., Porfiri, P. and Polakis, P. 1997 Stabilization of  $\beta$ -catenin by genetic defects in melanoma cell lines. *Science* 275, 1790-1792.

51. Polakis, P. The adenomatous polyposis coli (APC) tumor suppressor. 1997 *Biochem. Biophys. Acta*, 1332, F127-F147.

52. Rubinfeld, B., Albert, I., Porfiri, E., Munemitsu, S., and Polakis, P. 1997 Loss of  $\beta$ -catenin regulation by the APC tumor suppressor protein correlates with loss of structure due to common somatic mutations of the gene. *Cancer Res.* 57, 4624-4630.

53. Porfiri, E., Rubinfeld, B., Albert, I., Hovanes, K., Waterman, M., and Polakis, P. 1997 Induction of a  $\beta$ -catenin-LEF-1 complex by wnt-1 and transforming mutants of  $\beta$ -catenin. *Oncogene* 15, 2833-2839.

54. Thomas JE, Smith M, Tonkinson JL, Rubinfeld B, and Polakis P., 1997 Induction of phosphorylation on BRCA1 during the cell cycle and after DNA damage. *Cell Growth Differ.* 8, 801-809.

55. Hart, M., de los Santos, R., Albert, I., Rubinfeld, B., and Polakis P., 1998 Down regulation of  $\beta$ -catenin by human Axin and its association with the adenomatous polyposis coli (APC) tumor suppressor,  $\beta$ -catenin and glycogen synthase kinase 3 $\beta$ . *Current Biology* 8, 573-581.

56. Polakis, P. 1998 The oncogenic activation of  $\beta$ -catenin. *Current Opinions in Genetics and Development* 9, 15-21

57. Matt Hart, Jean-Paul Concordet, Irina Lassot, Iris Albert, Rico del los Santos, Herve Durand, Christine Perret, Bonnee Rubinfeld, Florence Margottin, Richard Benarous and Paul Polakis. 1999 The F-box protein  $\beta$ -TrCP associates with phosphorylated  $\beta$ -catenin and regulates its activity in the cell. *Current Biology* 9, 207-10.

---

58. Howard C. Crawford, Barbara M. Fingleton, Bonnee Rubinfeld, Paul Polakis and Lynn M. Matrisian 1999 The metalloproteinase matrilysin is a target of  $\beta$ -catenin transactivation in intestinal tumours. *Oncogene* 18, 2883-91.

59. Meng J, Glick JL, Polakis P, Casey PJ. 1999 Functional interaction between G $\alpha$ (z) and Rap1GAP suggests a novel form of cellular cross-talk. *J Biol Chem.* 17, 36663-9

60. Vijayasurian Easwaran, Virginia Song, Paul Polakis and Steve Byers 1999 The ubiquitin-proteasome pathway and serine kinase activity modulate APC mediated regulation of  $\beta$ -catenin-LEF signaling. *J. Biol. Chem.* 274(23):16641-5.

61. Polakis P, Hart M and Rubinfeld B. 1999 Defects in the regulation of beta-catenin in colorectal cancer. *Adv Exp Med Biol.* 470, 23-32

62. Shen Z, Batzer A, Koehler JA, Polakis P, Schlessinger J, Lydon NB, Moran MF. 1999 Evidence for SH3 domain directed binding and phosphorylation of Sam68 by Src. *Oncogene.* 18, 4647-53

64. Thomas GM, Frame S, Goedert M, Nathke I, Polakis P, Cohen P. 1999 A GSK3- binding peptide from FRAT1 selectively inhibits the GSK3-catalysed phosphorylation of axin and beta-catenin. *FEBS Lett.* 458, 247-51.

65. Peifer M, Polakis P. 2000 Wnt signaling in oncogenesis and embryogenesis--a look outside the nucleus. *Science* 287,1606-9.

66. Polakis P. 2000 Wnt signaling and cancer. *Genes Dev*;14, 1837-1851.

67. Spink KE, Polakis P, Weis WI 2000 Structural basis of the Axin-adenomatous polyposis coli interaction. *EMBO J* 19, 2270-2279.

68. Szeto, W., Jiang, W., Tice, D.A., Rubinfeld, B., Hollingshead, P.G., Fong, S.E., Dugger, D.L., Pham, T., Yansura, D.E., Wong, T.A., Grimaldi, J.C., Corpuz, R.T., Singh J.S., Frantz, G.D., Devaux, B., Crowley, C.W., Schwall, R.H., Eberhard,

D.A.,

Rastelli, L., Polakis, P. and Pennica, D. 2001 Overexpression of the Retinoic Acid-Responsive Gene Stra6 in Human Cancers and its Synergistic Induction by Wnt-1 and Retinoic Acid. *Cancer Res* 61, 4197-4204.

69. Rubinfeld B, Tice DA, Polakis P. 2001 Axin dependent phosphorylation of the adenomatous polyposis coli protein mediated by casein kinase 1 epsilon. *J Biol Chem* 276, 39037-39045.

---

70. Polakis P. 2001 More than one way to skin a catenin. *Cell* 2001 105, 563-566.

71. Tice DA, Soloviev I, Polakis P. 2002 Activation of the Wnt Pathway Interferes with Serum Response Element-driven Transcription of Immediate Early Genes. *J Biol. Chem.* 277, 6118-6123.

72. Tice DA, Szeto W, Soloviev I, Rubinfeld B, Fong SE, Dugger DL, Winer J,



Williams PM, Wieand D, Smith V, Schwall RH, Pennica D, Polakis P. 2002 Synergistic activation of tumor antigens by wnt-1 signaling and retinoic acid revealed by gene expression profiling. *J Biol Chem.* 277,14329-14335.

73. Polakis, P. 2002 Casein kinase I: A wnt'er of disconnect. *Curr. Biol.* 12, R499.

74. Mao, W., Luis, E., Ross, S., Silva, J., Tan, C., Crowley, C., Chui, C., Franz, G., Senter, P., Koeppen, H., Polakis, P. 2004 EphB2 as a therapeutic antibody drug target for the treatment of colorectal cancer. *Cancer Res.* 64, 781-788.

75. Shibamoto, S., Winer, J., Williams, M., Polakis, P. 2003 A Blockade in Wnt signaling is activated following the differentiation of F9 teratocarcinoma cells. *Exp. Cell Res.* 29211-20.

76. Zhang Y, Eberhard DA, Frantz GD, Dowd P, Wu TD, Zhou Y, Watanabe C, Luoh SM, Polakis P, Hillan KJ, Wood WI, Zhang Z. 2004 GEPIS—quantitative gene expression profiling in normal and cancer tissues. *Bioinformatics*, April 8.

---

## EXHIBIT B

	tumor mRNA	tumor IHC
UNQ2525	+	+
UNQ2378	+	+
UNQ972	+	-
UNQ97671	+	+
UNQ2964	+	+
UNQ323	+	+
UNQ1655	+	+
UNQ2333	+	+
UNQ9638	+	+
UNQ8209	+	+
UNQ6507	+	+
UNQ8196	+	+
UNQ9109	+	+
UNQ100	+	+
UNQ178	+	+
UNQ1477	+	+
UNQ1839	+	+
UNQ2079	+	+
UNQ8782	+	+
UNQ9646	+	-
UNQ111	+	+
UNQ3079	+	+
UNQ8175	+	+
UNQ9509	+	+
UNQ10978	+	-
UNQ2103	+	+
UNQ1563	+	+
UNQ16188	+	+
UNQ13589	+	+
UNQ1078	+	+
UNQ879	+	+

## Review

Paul A. Haynes  
Steven P. Gygi  
Daniel Figeys  
Ruedi Aebersold

Department of Molecular  
Biotechnology, University of  
Washington, Seattle, WA, USA

## Proteome analysis: Biological assay or data archive?

In this review we examine the current state of proteome analysis. There are three main issues discussed: why it is necessary to study proteomes; how proteomes can be analyzed with current technology; and how proteome analysis can be used to enhance biological research. We conclude that proteome analysis is an essential tool in the understanding of regulated biological systems. Current technology, while still mostly limited to the more abundant proteins, enables the use of proteome analysis both to establish databases of proteins present, and to perform biological assays involving measurement of multiple variables. We believe that the utility of proteome analysis in future biological research will continue to be enhanced by further improvements in analytical technology.

### Contents

1	Introduction .....	1862
2	Rationale for proteome analysis .....	1862
2.1	Correlation between mRNA and protein expression levels .....	1863
2.2	Proteins are dynamically modified and processed .....	1863
2.3	Proteomes are dynamic and reflect the state of a biological system .....	1863
3	Description and assessment of current proteome analysis technology .....	1863
3.1	Technical requirements of proteome technology .....	1863
3.2	2D electrophoresis – mass spectrometry: a common implementation of proteome analysis .....	1864
3.3	Protein identification by LC-MS/MS, capillary LC-MS/MS and CE-MS/MS .....	1865
3.3.1	LC-MS/MS .....	1865
3.3.2	Capillary LC-MS .....	1865
3.3.3	CE-MS/MS .....	1865
3.4	Assessment of 2-DE-MS proteome technology .....	1866
4	Utility of proteome analysis for biological research .....	1868
4.1	The proteome as a database .....	1868
4.2	The proteome as a biological assay .....	1868
5	Concluding remarks .....	1870
6	References .....	1870

### 1 Introduction

A proteome has been defined as the protein complement expressed by the genome of an organism, or, in multicellular organisms, as the protein complement expressed by a tissue or differentiated cell [1]. In the most common implementation of proteome analysis the proteins extracted from the cell or tissue analyzed are separated by high

resolution two-dimensional gel electrophoresis (2-DE), detected in the gel and identified by their amino acid sequence. The ease, sensitivity and speed with which gel-separated proteins can be identified by the use of recently developed mass spectrometric techniques have dramatically increased the interest in proteome technology. One of the most attractive features of such analyses is that complex biological systems can potentially be studied in their entirety, rather than as a multitude of individual components. This makes it far easier to uncover the many complex, and often obscure, relationships between mature gene products in cells. Large-scale proteome characterization projects have been undertaken for a number of different organisms and cell types. Microbial proteome projects currently in progress include, for example: *Saccharomyces cerevisiae* [2], *Salmonella enterica* [3], *Spiroplasma melliferum* [4], *Mycobacterium tuberculosis* [5], *Ochrobactrum anthropi* [6], *Haemophilus influenzae* [7], *Synechocystis* spp. [8], *Escherichia coli* [9], *Rhizobium leguminosarum* [10], and *Dictyostelium discoideum* [11]. Proteome projects underway for tissues of more complex organisms include those for: human bladder squamous cell carcinomas [12], human liver [13], human plasma [13], human keratinocytes [12], human fibroblasts [12], mouse kidney [12], and rat serum [14]. In this manuscript we critically assess the concept of proteome analysis and the technical feasibility of establishing complete proteome maps, and discuss ways in which proteome analysis and biological research intersect.

### 2 Rationale for proteome analysis

The dramatic growth in both the number of genome projects and the speed with which genome sequences are being determined has generated huge amounts of sequence information, for some species even complete genomic sequences ([15–17]). The description of the state of a biological system by the quantitative measurement of system components has long been a primary objective in molecular biology. With recent technical advances including the development of differential display-PCR [18], cDNA microarray and DNA chip technology [19, 20] and serial analysis of gene expression (SAGE) [21, 22], it is now feasible to establish global and quantitative mRNA expression maps of cells and tissues, in which the sequence of all the genes is known, at a speed and sensitivity which is not matched by current

Correspondence: Professor Ruedi Aebersold, Department of Molecular Biotechnology, University of Washington, Box 357730, Seattle, WA, 98195, USA (Tel: +206-685-4235; Fax: +206-685-6392; E-mail: ruedi@u.washington.edu)

Abbreviations: CID, collision-induced dissociation; MS/MS, tandem mass spectrometry; SAGE, serial analysis of gene expression

Keywords: Proteome / Two-dimensional polyacrylamide gel electrophoresis / Tandem mass spectrometry

protein analysis technology. Given the long-standing paradigm in biology that DNA synthesizes RNA which synthesizes protein, and the ability to rapidly establish global, quantitative mRNA expression maps, the questions which arise are why technically complex proteome projects should be undertaken and what specific types of information could be expected from proteome projects which cannot be obtained from genomic and transcript profiling projects. We see three main reasons for proteome analysis to become an essential component in the comprehensive analysis of biological systems. (i) Protein expression levels are not predictable from the mRNA expression levels, (ii) proteins are dynamically modified and processed in ways which are not necessarily apparent from the gene sequence, and (iii) proteomes are dynamic and reflect the state of a biological system.

## 2.1 Correlation between mRNA and protein expression levels

Interpretations of quantitative mRNA expression profiles frequently implicitly or explicitly assume that for specific genes the transcript levels are indicative of the levels of protein expression. As part of an ongoing study in our laboratory, we have determined the correlation of expression at the mRNA and protein levels for a population of selected genes in the yeast *Saccharomyces cerevisiae* growing at mid-log phase (S. P. Gygi *et al.*, submitted for publication). mRNA expression levels were calculated from published SAGE frequency tables [22]. Protein expression levels were quantified by metabolic radiolabeling of the yeast proteins, liquid scintillation counting of the protein spots separated by high resolution 2-DE and mass spectrometric identification of the protein(s) migrating to each spot. The selected 80 samples constitute a relatively homogeneous group with respect to predicted half-life and expression level of the protein products. Thus far, we have found a general trend but no strong correlation between protein and transcript levels (Fig. 1). For some genes studied equivalent mRNA transcript levels translated into protein abundances which varied by more than 50-fold. Similarly, equivalent steady-state protein expression levels were maintained by transcript levels varying by as much as 40-fold (S. P. Gygi *et al.*, submitted). These results suggest that even for a population of genes predicted to be relatively homogeneous with respect to protein half-life and gene expression, the protein levels cannot be accurately predicted from the level of the corresponding mRNA transcript.

## 2.2 Proteins are dynamically modified and processed

In the mature, biologically active form many proteins are post-translationally modified by glycosylation, phosphorylation, prenylation, acylation, ubiquitination or one or more of many other modifications [23] and many proteins are only functional if specifically associated or complexed with other molecules, including DNA, RNA, proteins and organic and inorganic cofactors. Frequently, modifications are dynamic and reversible and may alter the precise three-dimensional structure and the state of activity of a protein. Collectively, the state of modification of the proteins which constitute a biological system

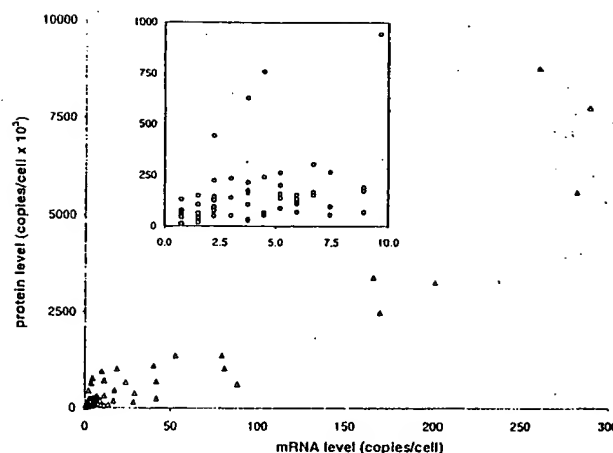


Figure 1. Correlation between mRNA and protein levels in yeast cells. For a selected population of 80 genes, protein levels were measured by  $^{35}\text{S}$ -radiolabeling and mRNA levels were calculated from published SAGE tables. Inset: expanded view of the low abundance region. For more experimental details, also see Figs. 5 and 6, (S. P. Gygi *et al.*, submitted).

are important indicators for the state of the system. The type of protein modification and the sites modified at a specific cellular state can usually not be determined from the gene sequence alone.

## 2.3 Proteomes are dynamic and reflect the state of a biological system

A single genome can give rise to many qualitatively and quantitatively different proteomes. Specific stages of the cell cycle and states of differentiation, responses to growth and nutrient conditions, temperature and stress, and pathological conditions represent cellular states which are characterized by significantly different proteomes. The proteome, in principle, also reflects events that are under translational and post-translational control. It is therefore expected that proteomics will be able to provide the most precise and detailed molecular description of the state of a cell or tissue, provided that the external conditions defining the state are carefully determined. In answer to the question of whether the study of proteomes is necessary for the analysis of biomolecular systems, it is evident that the analysis of mature protein products in cells is essential as there are numerous levels of control of protein synthesis, degradation, processing and modification, which are only apparent by direct protein analysis.

## 3 Description and assessment of current proteome analysis technology

### 3.1 Technical requirements of proteome technology

In biological systems the level of expression as well as the states of modification, processing and macro-molecular association of proteins are controlled and modulated depending on the state of the system. Comprehensive analysis of the identity, quantity and state of modification of proteins therefore requires the detection and

quantitation of the proteins which constitute the system, and analysis of differentially processed forms. There are a number of inherent difficulties in protein analysis which complicate these tasks. First, proteins cannot be amplified. It is possible to produce large amounts of a particular protein by over-expression in specific cell systems. However, since many proteins are dynamically post-translationally modified, they cannot be easily amplified in the form in which they finally function in the biological system. It is frequently difficult to purify from the native source sufficient amounts of a protein for analysis. From a technological point of view this translates into the need for high sensitivity analytical techniques. Second, many proteins are modified and processed post-translationally. Therefore, in addition to the protein identity, the structural basis for differentially modified isoforms also needs to be determined. The distribution of a constant amount of protein over several differentially modified isoforms further reduces the amount of each species available for analysis. The complexity and dynamics of post-translational protein editing thus significantly complicates proteome studies. Third, proteins vary dramatically with respect to their solubility in commonly used solvents. There are few, if any, solvent conditions in which all proteins are soluble and which are also compatible with protein analysis. This makes the development of protein purification methods particularly difficult since both protein purification and solubility have to be achieved under the same conditions. Detergents, in particular sodium dodecyl sulfate (SDS), are frequently added to aqueous solvents to maintain protein solubility. The compatibility with SDS is a big advantage of SDS polyacrylamide gel electrophoresis (SDS-PAGE) over other protein separation techniques. Thus, SDS-PAGE and two-dimensional gel electrophoresis, which also uses SDS and other detergents, are the most general and preferred methods for the purification of small amounts of proteins, provided that activity does not necessarily need to be maintained. Lastly, the number of proteins in a given cell system is typically in the thousands. Any attempt to identify and categorize all of these must use methods which are as rapid as possible to allow completion of the project within a reasonable time frame. Therefore, a successful, general proteomics technology requires high sensitivity, high throughput, the ability to differentiate differentially modified proteins, and the ability to quantitatively display and analyze all the proteins present in a sample.

### 3.2 2-D electrophoresis – mass spectrometry: a common implementation of proteome analysis

The most common currently used implementation of proteome analysis technology is based on the separation of proteins by two-dimensional (IEF/SDS-PAGE) gel electrophoresis and their subsequent identification and analysis by mass spectrometry (MS) or tandem mass spectrometry (MS/MS). In 2-DE, proteins are first separated by isoelectric focusing (IEF) and then by SDS-PAGE, in the second, perpendicular dimension. Separated proteins are visualized at high sensitivity by staining or autoradiography, producing two-dimensional arrays of proteins. 2-DE gels are, at present, the most commonly used means of global display of proteins in complex

samples. The separation of thousands of proteins has been achieved in a single gel [24, 25] and differentially modified proteins are frequently separated. Due to the compatibility of 2-DE with high concentrations of detergents, protein denaturants and other additives promoting protein solubility, the technique is widely used.

The second step of this type of proteome analysis is the identification and analysis of separated proteins. Individual proteins from polyacrylamide gels have traditionally been identified using *N*-terminal sequencing [26, 27], internal peptide sequencing [28, 29], immunoblotting or comigration with known proteins [30]. The recent dramatic growth of large-scale genomic and expressed sequence tag (EST) sequence databases has resulted in a fundamental change in the way proteins are identified by their amino acid sequence. Rather than by the traditional methods described above, protein sequences are now frequently determined by correlating mass spectral or tandem mass spectral data of peptides derived from proteins, with the information contained in sequence databases [31–33].

There are a number of alternative approaches to proteome analysis currently under development. There is considerable interest in developing a proteome analysis strategy which bypasses 2-DE altogether, because it is considered a relatively slow and tedious process, and because of perceived difficulties in extracting proteins from the gel matrix for analysis. However, 2-DE as a starting point for proteome analysis has many advantages compared to other techniques available today. The most significant strengths of the 2-DE-MS approach include the relatively uniform behavior of proteins in gels, the ability to quantify spots and the high resolution and simultaneous display of hundreds to thousands of proteins within a reasonable time frame.

A schematic diagram of a typical procedure of the identification of gel-separated proteins is shown in Fig. 2. Protein spots detected in the gel are enzymatically or chemically fragmented and the peptide fragments are isolated for analysis, as already indicated, most frequently by MS or MS/MS. There are numerous protocols for the generation of peptide fragments from gel-separated proteins. They can be grouped into two categories, digestion in the gel slice [28, 34] or digestion after electrotransfer out of the gel onto a suitable membrane ([29, 35–37] and reviewed in [38]). In most instances either technique is applicable and yields good results. The analysis of MS or MS/MS data is an important step in the whole process because MS instruments can generate an enormous amount of information which cannot easily be managed manually. Recently, a number of groups have developed software systems dedicated to the use of peptide MS and MS/MS spectra for the identification of proteins. Proteins are identified by correlating the information contained in the MS spectra of protein digests or MS/MS spectra of individual peptides with data contained in DNA or protein sequence databases.

The systems we are currently using in our laboratory are based on the separation of the peptides contained in protein digests by narrow bore or capillary liquid chromatog-

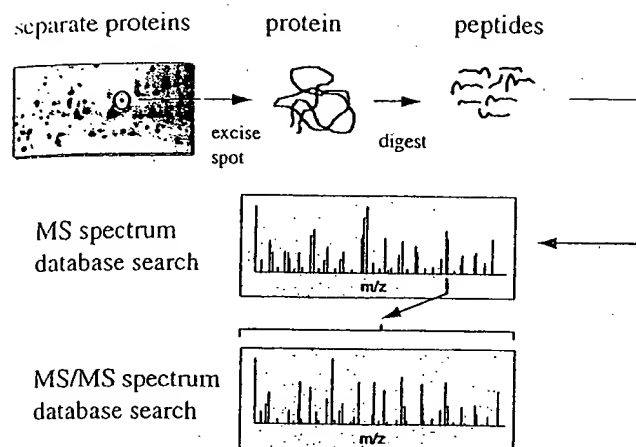


Figure 2. Schematic diagram of a procedure for identification of gel-separated proteins. Peptides can either be separated by a technique such as LC or CE, or infused as a mixture and sorted in the MS. Database searching can either be performed on peptide masses from an MS spectrum, peptide fragment masses from CID spectra of peptides, or a combination of both.

raphy [39, 40] or capillary electrophoresis [41], the analysis of the separated peptides by electrospray ionization (ESI) MS/MS, and the correlation of the generated peptide spectra with sequence databases using the SEQUEST program developed at the University of Washington [32, 33]. The system automatically performs the following operations: a particular peptide ion characterized by its mass-to-charge ratio is selected in the MS out of all the peptide ions present in the system at a particular time; the selected peptide ion is collided in a collision cell with argon (collision-induced dissociation, CID) and the masses of the resulting fragment ions are determined in the second sector of the tandem MS; this experimentally determined CID spectrum is then correlated with the CID spectra predicted from all the peptides in a sequence database which have essentially the same mass as the peptide selected for CID; this correlation matches the isolated peptide with a sequence segment in a database and thus identifies the protein from which the peptide was derived. There are a number of alternative programs which use peptide CID spectra for protein identification, but we use the SEQUEST system because it is currently the most highly automated program and has proven to be successful, versatile and robust.

### 3.3 Protein identification by LC-MS/MS, capillary LC-MS/MS and CE-MS/MS

It has been demonstrated repeatedly that MS has a very high intrinsic sensitivity. For the routine analysis of gel-separated proteins at high sensitivity, the most significant challenge is the handling of small amounts of sample. The crux of the problem is the extraction and transfer of peptide mixtures generated by the digestion of low nanogram amounts of protein, from gels into the MS/MS system without significant loss of sample or introduction of unwanted contaminants. We employ three different systems for introducing gel-purified samples into an MS, depending on the level of sensitivity

required. As an approximate guideline, for samples containing tens of picomoles of peptides, LC-MS/MS is most appropriate; for samples containing low picomole amounts to high femtomole amounts we use capillary LC-MS/MS; and for samples containing femtomoles or less, CE-MS/MS is the method of choice.

#### 3.3.1 LC-MS/MS

The coupling of an MS to an HPLC system using a 0.5 mm diameter or bigger reverse phase (RP) column has been described in detail [42]. This system has several advantages if a large number of samples are to be analyzed and all are available in sufficient quantity. The LC-MS and database searching program can be run in a fully automated mode using an autosampler, thus maximizing sample throughput and minimizing the need for operator interference. The relatively large column is tolerant of high levels of impurities from either gel preparation or sample matrix. Lastly, if configured with a flow-splitter and micro-sprayer [40], analyses can be performed on a small fraction of the sample (less than 5%) while the remainder of the sample is recovered in very pure solvents. This latter feature is particularly useful when an orthogonal technique is also used to analyze peptide fractions, such as scintillation of an introduced radiolabel, and this data can be correlated with peptides identified by CID spectra.

#### 3.3.2 Capillary LC-MS

An increase of sensitivity of approximately tenfold can be achieved by using a capillary LC system with a 100  $\mu$ m ID column rather than a 0.5 mm ID column as referred to above. Since very low flow rates are required for such columns, most reports have used a precolumn flow splitting system for producing solvent gradients. We have recently described the design and construction of a novel gradient mixing system which enables the formation of reproducible gradients at very low flow rates (low nL/min) without the need for flow splitting (A. Ducret *et al.*, submitted for publication). Using this capillary LC-MS/MS system we were able to identify gel-separated proteins if low picomole to high femtomole amounts were loaded onto the gel [40]. This system is as yet not automated and, like all capillary LC systems, is prone to blockage of the columns by microparticulates when analyzing gel-separated proteins.

#### 3.3.3 CE-MS/MS

The highest level of sensitivity for analyzing gel-separated proteins can be achieved by using capillary electrophoresis – mass spectrometry (CE-MS). We have described in the past a solid-phase extraction capillary electrophoresis (SPE-CE) system which was used with triple quadrupole and ion trap ESI-MS/MS systems for the identification of proteins at the low femtomole to sub-femtomole sensitivity level [43, 44]. While this system is highly sensitive, its operation is labor-intensive and its operation has not been automated. In order to devise an analytical system with both the sensitivity of a CE and the level of automation of LC, we have constructed

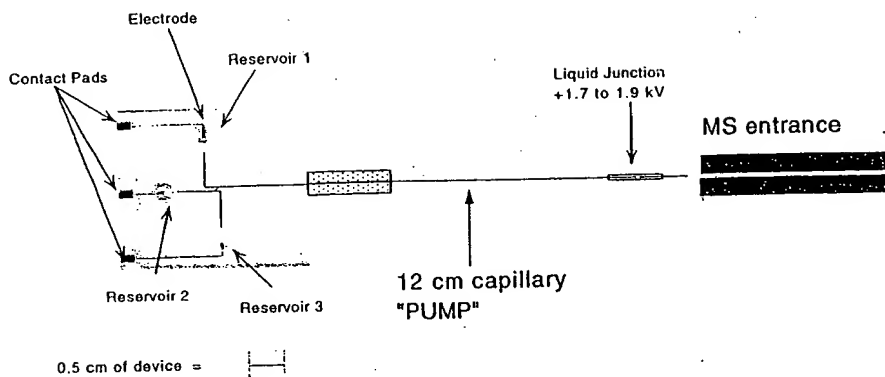


Figure 3. Schematic illustration of a microfabricated analytical system for CE, consisting of a micromachined device, coated capillary electroosmotic pump, and microelectrospray interface. The dimensions of the channels and reservoir are as indicated in the text. The channels on the device were graphically enhanced to make them more visible. Reproduced from [45], with permission.

microfabricated devices for the introduction of samples into ESI-MS for high-sensitivity peptide analysis.

The basic device is a piece of glass into which channels of 10–30  $\mu\text{m}$  in depth and 50–70  $\mu\text{m}$  in diameter are etched by using photolithography/etching techniques similar to the ones used in the semiconductor industry. (A simple device is shown in Fig. 3). The channels are connected to an external high voltage power supply [45]. Samples are manipulated on the device and off the device to the MS by applying different potentials to the reservoirs. This creates a solvent flow by electroosmotic pumping which can be redirected by changing the position of the electrode. Therefore, without the need for valves or gates and without any external pumping, the flow can be redirected by simply switching the position of the electrodes on the device. The direction and rate of the flow can be modulated by the size and the polarity of the electric field applied and also by the charge state of the surface.

The type of data generated by the system is illustrated in Fig. 4, which shows the mass spectrum of a peptide sample representing the tryptic digest of carbonic anhydrase at 290 fmol/ $\mu\text{L}$ . Each numbered peak indicates a peptide successfully identified as being derived from carbonic an-

hydrase. Some of the unassigned signals may be chemical or peptide contaminants. The MS is programmed to automatically select each peak and subject the peptide to CID. The resulting CID spectra are then used to identify the protein by correlation with sequence databases. Therefore, this system allows us to concurrently apply a number of protein digests onto the device, to sequentially mobilize the samples, to automatically generate CID spectra of selected peptide ions and to search sequence databases for protein identification. These steps are performed automatically without the need for user input and proteins can be identified at very low femtomole level sensitivity at a rate of approximately one protein per 15 min.

### 3.4 Assessment of 2-DE-MS proteome technology

Using a combination of the analytical techniques described above we have identified the 80 protein spots indicated in Fig. 5. The protein pattern was generated by separating a total of 40 microgram of protein contained in a total cell lysate of the yeast strain YPH499 by high resolution 2-DE and silver staining of the separated proteins. To estimate how far this type of proteome analysis can penetrate towards the identification of low abundance proteins, we have calculated the codon bias of the genes encoding the respective proteins. Codon bias is a

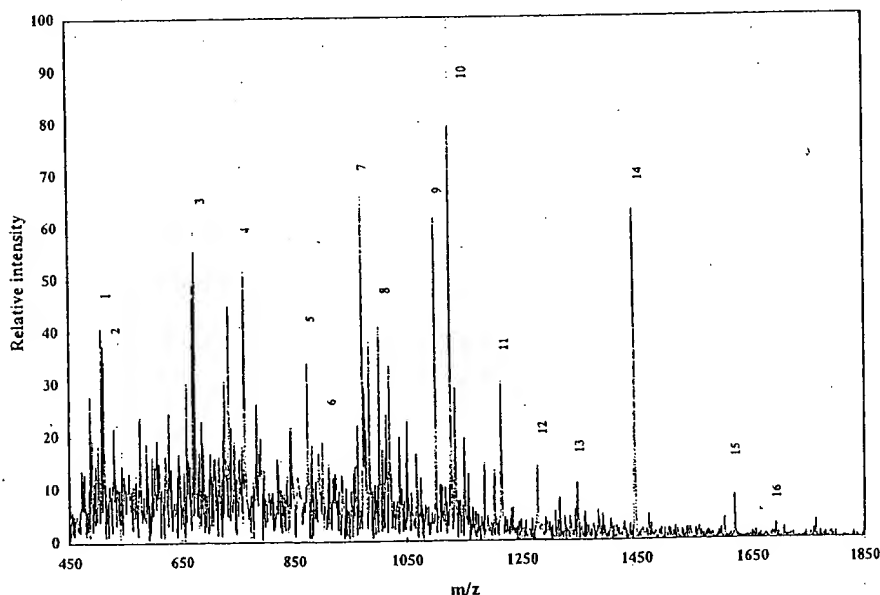


Figure 4. MS spectrum of a tryptic digest of carbonic anhydrase using the microfabricated system shown in Fig. 3. 290 fmol/ $\mu\text{L}$  of carbonic anhydrase tryptic digest was infused into a Finnigan LCQ ion trap MS. Each peak was selected for CID, and those which were identified as containing peptides derived from carbonic anhydrase are numbered. Reproduced from [45], with permission.



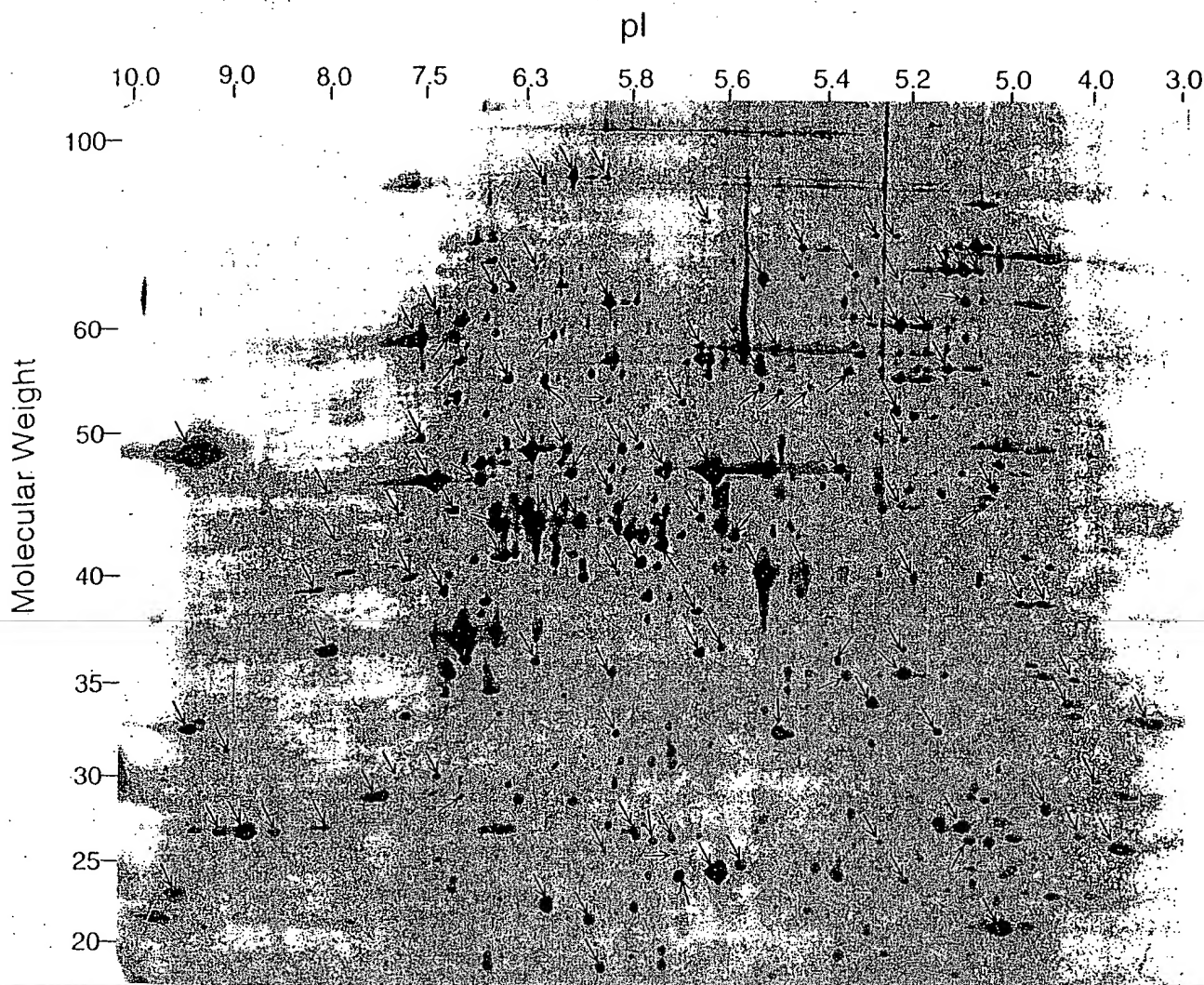


Figure 5. 2-DE separation of a lysate of yeast cells, with identified proteins highlighted. The first dimension of separation was an IPG from pH 3–10, and the second dimension was a 10%T SDS-PAGE gel. Proteins were visualized by silver staining. Further details of experimental procedures are included in S. P. Gygi *et al.* (submitted).

calculated measure of the degree of redundancy of triplet DNA codons used to produce each amino acid in a particular gene sequence. It has been shown to be a useful indicator of the level of the protein product of a particular gene sequence present in a cell [46]. The general rule which applies is that the higher the value of the codon bias calculated for a gene, the more abundant the protein product of that gene becomes. The calculated codon bias values corresponding to the proteins identified in Fig. 5 are shown in Fig. 6b. Nearly all of the proteins identified (> 95%) have codon bias values of > 0.2, indicating they are highly abundant in cells. In contrast, codon bias values calculated for the entire yeast genome (Fig. 6a) show that the majority of proteins present in the proteome have a codon bias of < 0.2 and are thus of low abundance.

This finding is of considerable importance in our assessment of the current status of proteome analysis technology. It is clear that even using highly sensitive analytical techniques, we are only able to visualize and identify the

more abundant proteins. Since many important regulatory proteins are present only at low abundance, these would not be amenable to analysis using such techniques. This situation would be exacerbated in the analysis of proteomes containing many more proteins than the approximately 6000 gene products present in yeast cells [16]. In the analysis of, for example, the proteome of any human cells, there are potentially 50 000–100 000 gene products [47]. Inherent limitations on the amount of protein that can be loaded on 2-DE, and the number of components that can be resolved, indicate that only the most highly abundant fraction of the many gene products could be successfully analyzed. One approach that has been employed to circumvent these limitations is the use of very narrow range immobilized pH gradient strips for the first-dimension separation of 2-DE [48]. Since only those proteins which focus within the narrow range will enter the second dimension of separation, a much higher sample loading within the desired range is possible. This, in turn, can lead to the visualization and identification of less abundant proteins.



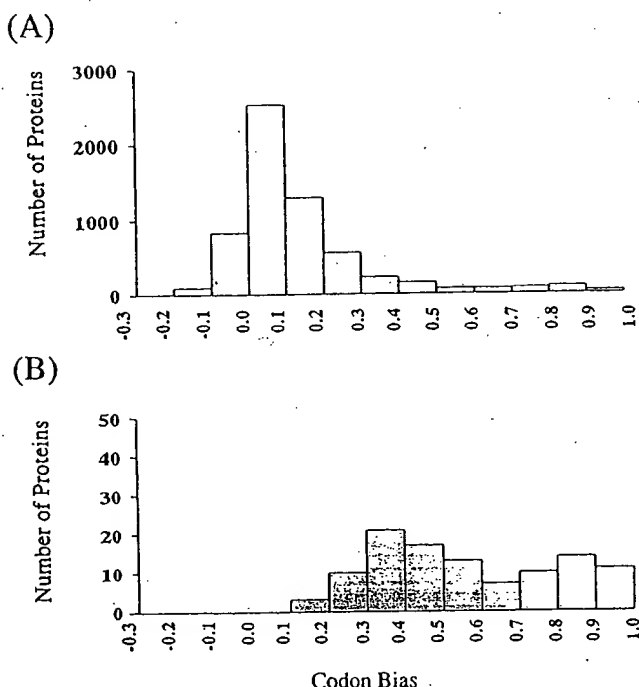


Figure 6. Calculated codon bias values for yeast proteins. (A) Distribution of calculated values for the entire yeast proteome. (B) Distribution of calculated values for the subset of 80 identified proteins also shown in Figs. 1 and 5. Further details of experimental procedures are included in S. P. Gygi *et al.* (submitted).

#### 4 Utility of proteome analysis for biological research

For the success of proteomics as a mainstream approach to the analysis of biological systems it is essential to define how proteome analysis and biological research projects intersect. Without a clear plan for the implementation of proteome-type approaches into biological research projects the full impact of the technology can not be realized. The literature indicates that proteome analysis is used both as a database/data archive, and as a biological assay or biological research tool.

##### 4.1 The proteome as a database

The use of proteomics as a database or data archive essentially entails an attempt to identify all the proteins in a cell or species and to annotate each protein with the known biological information that is relevant for each protein. The level of annotation can, of course, be extensive. The most common implementation of this idea is the separation of proteins by high resolution 2-DE, the identification of each detected protein spot and the annotation of the protein spots in a 2-DE gel database format. This approach is complicated by the fact that it is difficult to precisely define a proteome and to decide which proteome should be represented in the database. In contrast to the genome of a species, which is essentially static, the proteome is highly dynamic. Processes such as differentiation, cell activation and disease can all significantly change the proteome of a species. This is illustrated in Fig. 7. The figure shows two high-resolu-

tion 2-DE maps of proteins isolated from rat serum. Fig. 7A is from the serum of normal rats, while Fig. 7B is from the serum of rats in acute-phase serum after prior treatment with an inflammation-causing agent [49]. It is obvious that the protein patterns are significantly different in several areas, raising the question of exactly which proteome is being described.

Therefore, a comprehensive proteome database of a species or cell type needs to contain all of the parameters which describe the state and the type of the cells from which the proteins were extracted as well as the software tools to search the database with queries which reflect the dynamics of biological systems. A comprehensive proteome database should be capable of quantitatively describing the fate of each protein if specific systems and pathways are activated in the cell. Specifically, the quantity, the degree of modification, the subcellular location and the nature of molecules specifically interacting with a protein as well as the rate of change of these variables should be described. Using these admittedly stringent criteria, there is currently no complete proteome database. A number of such databases are, however, in the process of being constructed. The most advanced among them, in our opinion, are the yeast protein database YPD [50] (accessible at <http://www.ypd.com>) and the human 2D-PAGE databases of the Danish Centre for Human Genome Research [12] (accessible at <http://biobase.dk/cgi-bin/celis>). While neither can be considered complete as not all of the potential gene products are identified, both contain extensive annotation of supplemental information for many of the spots which are positively identified in reference samples.

##### 4.2 The proteome as a biological assay

The use of proteome analysis as a biological assay or research tool represents an alternative approach to integrating biology with proteomics. To investigate the state of a system, samples are subjected to a specific process that allows the quantitative or qualitative measurement of some of the variables which describe the system. In typical biochemical assays one variable (*e.g.*, enzyme activity) of a single component (*e.g.*, a particular enzyme) is measured. Using proteomics as an assay, multiple variables (*e.g.*, expression level, rate of synthesis, phosphorylation state, etc.) are measured concurrently on many (ideally all) of the proteins in a sample. The use of proteomics as an assay is a less far-reaching proposition than the construction of a comprehensive proteome database. It does, however, represent a pragmatic approach which can be adapted to investigate specific systems and pathways, as long as the interpretation of the results takes into account that with current technology not all of the variables which describe the system can be observed (see Section 3.4).

A common implementation of proteome analysis as a biological assay is when a 2-DE protein pattern generated from the analysis of an experimental sample is compared to an array of reference patterns representing different states of the system under investigation. The state of the experimental system at the time the sample was generated is therefore determined by the quantita-

tive comparative analysis of hundreds to a few thousand proteins. Comparative analysis of the 2-DE patterns furthermore highlights quantitative and qualitative differences in the protein profiles which correlate with the state of the system. For this type of analysis it is not essential that all the proteins are identified or even visu-

alized, although the results become more informative as more proteins are compared. It is obvious, however, that the possibility to identify any protein deemed characteristic for a particular state dramatically enhances this approach by opening up new avenues for experimentation.

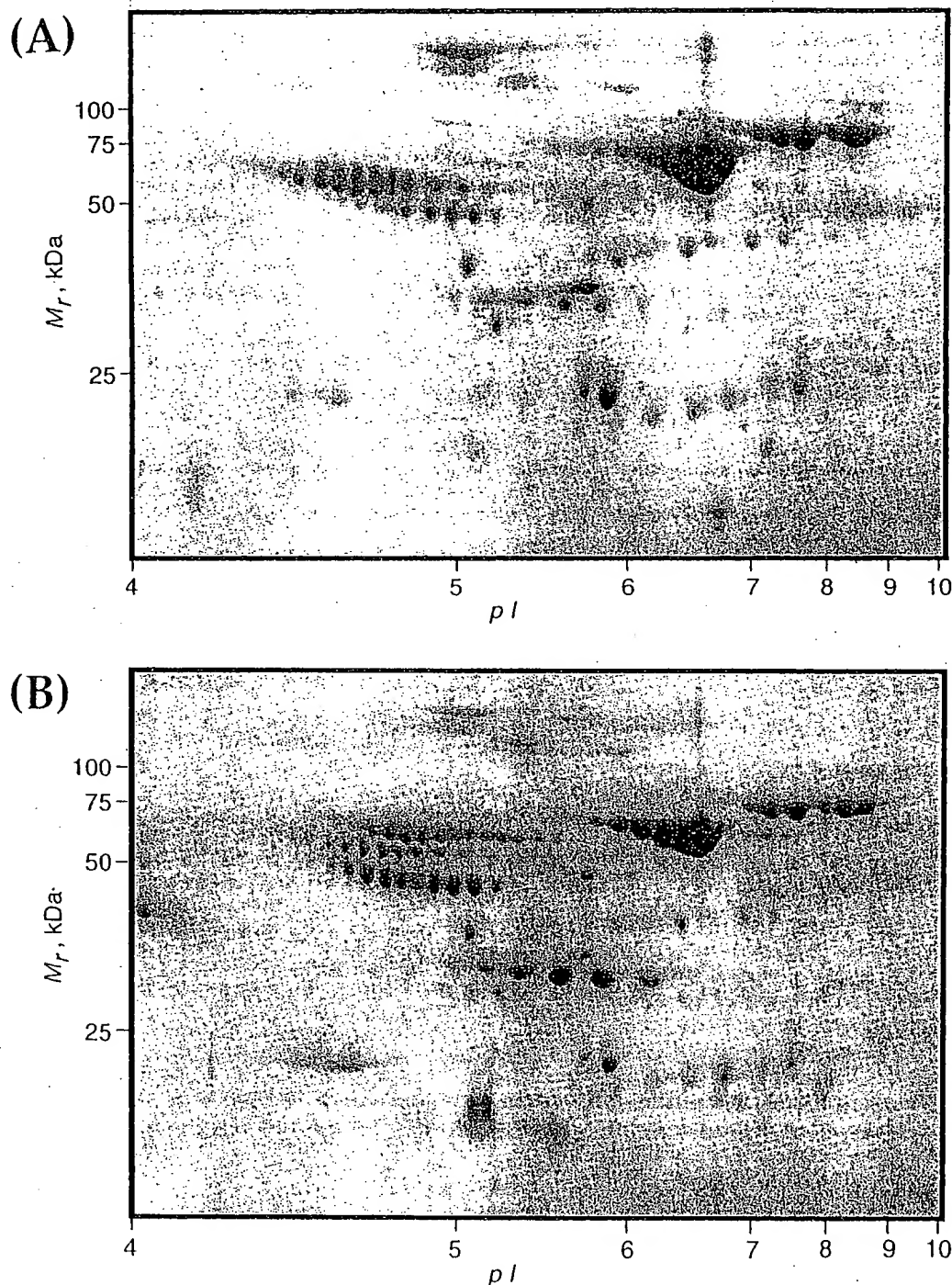


Figure 7. High resolution 2-DE map of proteins isolated from rat serum with or without prior exposure to an inflammation-causing agent. (A) normal rat serum, (B) acute-phase serum from rats which had previously been exposed to an inflammation-causing agent. The first dimension of separation is an IPG from pH 4–10, and the second dimension is a 7.5–17.5%T gradient SDS-PAGE gel. Proteins were visualized by staining with amido black. Further details of experimental procedures are included in [14, 49].

Proteome analysis as a biological assay has been successfully used in the field of toxicology, to characterize disease states or to study differential activation of cells. The approach is limited, of course, by the fact that only the visible protein spots are included in the assay, and it is well known that a substantial but far from complete fraction of cellular proteins are detected if a total cell lysate is separated by 2-DE. Proteins may not be detected in 2-DE gels because they are not abundant enough to be visualized by the detection method used, because they do not migrate within the boundaries (size, *pI*) resolved by the gel, because they are not soluble under the conditions used, or for other reasons.

A different way to use proteome analysis as a biological assay to define the state of a biological system is to take advantage of the wealth of information contained in 2-DE protein patterns. 2-DE is referred to as two-dimensional because of the electrophoretic mobility and the isoelectric points which define the position of each protein in a 2-DE pattern. In addition to the two dimensions used to generate the protein patterns, a number of additional data dimensions are contained in the protein patterns. Some of these dimensions such as protein expression level, phosphorylation state, subcellular location, association with other proteins, rate of synthesis or degradation indicate the activity state of a protein or a biological system. Comparative analysis of 2-DE protein patterns representing different states is therefore ideally suited for the detection, identification and analysis of suitable markers. Once again it must be emphasized that in this type of experiment only a fraction of the cellular proteins is analyzed. Since many regulatory proteins are of low abundance, this limitation is a concern, particularly in cases in which regulatory pathways are being investigated.

## 5 Concluding remarks

In this report we have addressed three main issues related to proteome analysis. First, we have discussed the rationale for studying proteomes. Second, we have assessed the technical feasibility of analyzing proteomes and described current proteome technology, and third, we have analyzed the utility of proteome analysis for biological research. It is apparent that proteome analysis is an essential tool in the analysis of biological systems. The multi-level control of protein synthesis and degradation in cells means that only the direct analysis of mature protein products can reveal their correct identities, their relevant state of modification and/or association and their amounts. Recently developed methods have enabled the identification of proteins at ever-increasing sensitivity levels and at a high level of automation of the analytical processes. A number of technical challenges, however, remain. While it is currently possible to identify essentially any protein spots that can be visualized by common staining methods, it is apparent that without prior enrichment only a relatively small and highly selected population of long-lived, highly expressed proteins is observed. There are many more proteins in a given cell which are not visualized by such methods. Frequently it is the low abundance proteins that execute key regulatory functions.

We have outlined the two principal ways proteome analysis is currently being used to intersect with biological research projects: the proteome as a database or data archive and proteome analysis as a biological assay. Both approaches have in common that at present they are conceptually and technically limited. Current proteome databases typically are limited to one cell type and one state of a cell and therefore do not account for the dynamics of biological systems. The use of proteome analysis as a biological assay can provide a wealth of information, but it is limited to the proteins detected and is therefore not truly proteome-wide. These limitations in proteomics are to a large extent a reflection of the fact that proteins in their fully processed form cannot easily be amplified and are therefore difficult to isolate in amounts sufficient for analysis or experimentation. The fact that to date no complete proteome has been described further attests to these difficulties. With continued rapid progress in protein analysis technology, however, we anticipate that the goal of complete proteome analysis will eventually become attainable.

*We would like to acknowledge the funding for our work from the National Science Foundation Science and Technology Center for Molecular Biotechnology and from the NIH. We thank Yvan Rochon and Bob Franza for providing the yeast gel shown and Elisabetta Gianazza for providing the rat serum gels shown.*

Received April 21, 1998

## 6 References

- [1] Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J.-C., Yan, J. X., Gooley, A. A., Hughes, G., Humphery-Smith, I., Williams, K. L., Hochstrasser, D. F., *Bio/Technology* 1996, 14, 61–65.
- [2] Hodges, P. E., Payne, W. E., Garrels, J. I., *Nucleic Acids Res.* 1998, 26, 68–72.
- [3] O'Connor, C. D., Farris, M., Fowler, R., Qi, S. Y., *Electrophoresis* 1997, 18, 1483–1490.
- [4] Cordwell, S. J., Basseal, D. J., Humphery-Smith, I., *Electrophoresis* 1997, 18, 1335–1346.
- [5] Urquhart, B. L., Atsalos, T. E., Roach, D., Basseal, D. J., Bjellqvist, B., Britton, W. L., Humphery-Smith, I., *Electrophoresis* 1997, 18, 1384–1392.
- [6] Wasinger, V. C., Bjellqvist, B., Humphery-Smith, I., *Electrophoresis* 1997, 18, 1373–1383.
- [7] Link, A. J., Hays, L. G., Carmack, E. B., Yates III, J. R., *Electrophoresis* 1997, 18, 1314–1334.
- [8] Sazuka, T., Ohara, O., *Electrophoresis* 1997, 18, 1252–1258.
- [9] VanBogelen, R. A., Abshire, K. Z., Moldover, B., Olson, E. R., Neidhardt, F. C., *Electrophoresis* 1997, 18, 1243–1251.
- [10] Guerreiro, N., Redmond, J. W., Rolfe, B. G., Djordjevic, M. A., *Mol. Plant Microbe Interact.* 1997, 10, 506–516.
- [11] Yan, J. X., Tonella, L., Sanchez, J.-C., Wilkins, M. R., Packer, N. H., Gooley, A. A., Hochstrasser, D. F., Williams, K. L., *Electrophoresis* 1997, 18, 491–497.
- [12] Celis, J., Gromov, P., Ostergaard M., Madsen, P., Honoré, B., Deigaard, K., Olsen, E., Vorum, H., Kristensen, D. B., Gromova, I., Haunso, A., Van Damme, J., Puype, M., Vandekerckhove, J., Rasmussen, H. H., *FEBS Lett.* 1996, 398, 129–134.
- [13] Appel, R. D., Sanchez, J.-C., Bairoch, A., Golaz, O., Miu, M., Vargas, J. R., Hochstrasser, D. F., *Electrophoresis* 1993, 14, 1232–1238.
- [14] Haynes, P., Miller, I., Aebersold, R., Gemeiner, M., Eberini, I., Lovati, R. M., Manzoni, C., Vignati, M., Gianazza, E., *Electrophoresis* 1998, 19, 1484–1492.

- [15] Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J.-F., Dougherty, B. A., Merrick, J. M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J. D., Scott, J., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J. M., Weidman, J. F., Phillips, C. A., Spriggs, T., Hedblom, E., Cotton, M. D., Utterback, T. R., Hanna, N. C., Nguyen, D. T., Saudek, D. M., Brandon, R. C., Fine, L. D., Fritchman, J. L., Fuhrmann, J. L., Geoghegan, N. S. M., Gnehm, C. L., McDonald, L. A., Small, K. V., Fraser, C. M., Smith, C. O., Venter, J. C., *Science* 1995, 269, 496-512.
- [16] Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., Oliver, S. G., *Science* 1996, 274, 546.
- [17] Fraser, C. M., Casjens, S., Huang, W. M., Sutton, G. G., Clayton, R., Lathigra, R., White, O., Ketchum, K. A., Dodson, R., Hickey, E. K., Gwinn, M., Dougherty, B., Tomb, J. F., Fleischmann, R. D., Richardson, D., Peterson, J., Kerlavage, A. R., Quackenbush, J., Salzberg, S., Hanson, M., van Vugt, R., Palmer, N., Adams, M. D., Gocayne, J., Weidman, J., Utterback, T., Watthey, T., McDonald, L., Artiach, P., Bowman, C., Garland, S., Fujii, C., Cotton, M. D., Horst, K., Roberts, K., Hatch, B., Smith, H. O., Venter, J. C., *Nature* 1997, 390, 580-586.
- [18] Liang, P., Pardee, A. B., *Science* 1992, 257, 967-971.
- [19] Lashkari, D. A., DeRisi, J. L., McCusker, J. H., Namath, A. F., Gentile, C., Hwang, S. Y., Brown, P. O., Davis, R. W., *Proc. Natl. Acad. Sci. USA* 1997, 94, 13057-13062.
- [20] Shalon, D., Smith, S. J., Brown, P. O., *Genome Res.* 1996, 6, 639-645.
- [21] Velculescu, V. E., Zhang, L., Vogelstein, B., Kinzler, K. W., *Science* 1995, 270, 484-487.
- [22] Velculescu, V. E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M. A., Bassett, D. E., Hieter, P., Vogelstein, B., Kinzler, K. W., *Cell* 1997, 88, 243-251.
- [23] Krishna, R. G., Wold, F., *Adv. Enzymol.* 1993, 67, 265-298.
- [24] Görg, A., Postel, W., Gunther, S., *Electrophoresis* 1988, 9, 531-546.
- [25] Klose, J., Kobalz, U., *Electrophoresis* 1995, 16, 1034-1059.
- [26] Matsudaira, P., *J. Biol. Chem.* 1987, 262, 10035-10038.
- [27] Aebersold, R. H., Teplow, D. B., Hood, L. E., Kent, S. B., *J. Biol. Chem.* 1986, 261, 4229-4238.
- [28] Rosenfeld, J., Capdevielle, J., Guillemot, J. C., Ferrara, P., *Anal. Biochem.* 1992, 203, 173-179.
- [29] Aebersold, R. H., Leavitt, J., Saavedra, R. A., Hood, L. E., Kent, S. B., *Proc. Natl. Acad. Sci. USA* 1987, 84, 6970-6974.
- [30] Honoré, B., Leffers, H., Madsen, P., Celis, J. E., *Eur. J. Biochem.* 1993, 218, 421-430.
- [31] Mann, M., Wilm, M., *Anal. Chem.* 1994, 66, 4390-4399.
- [32] Eng, J., McCormack, A. L., Yates III, J. R., *J. Amer. Mass Spectrom.* 1994, 5, 976-989.
- [33] Yates III, J. R., Eng, J. K., McCormack, A. L., Schieltz, D., *Anal. Chem.* 1995, 67, 1426-1436.
- [34] Shevchenko, A., Wilm, M., Vorm, O., Mann, M., *Anal. Chem.* 1996, 68, 850-858.
- [35] Hess, D., Covey, T. C., Winz, R., Brownsey, R. W., Aebersold, R., *Protein Sci.* 1993, 2, 1342-1351.
- [36] van Oostveen, I., Ducret, A., Aebersold, R., *Anal. Biochem.* 1997, 247, 310-318.
- [37] Lui, M., Tempst, P., Erdjument-Bromage, H., *Anal. Biochem.* 1996, 241, 156-166.
- [38] Patterson, S. D., Aebersold, R. A., *Electrophoresis* 1995, 16, 1791-1814.
- [39] Ducret, A., Foy, Brun, C., Bures, E. J., Marhaug, G., Husby, G. R. A., *Electrophoresis* 1996, 17, 866-876.
- [40] Haynes, P. A., Fripp, N., Aebersold, R., *Electrophoresis* 1998, 19, 939-945.
- [41] Figeys, D., Van Oostveen, I., Ducret, A., Aebersold, R., *Anal. Chem.* 1996, 68, 1822-1828.
- [42] Ducret, A., Van Oostveen, I., Eng, J. K., Yates III, J. R., Aebersold, R., *Protein Sci.* 1997, 7, 706-719.
- [43] Figeys, D., Ducret, A., Yates III, J. R., Aebersold, R., *Nature Biotech.* 1996, 14, 1579-1583.
- [44] Figeys, D., Aebersold, R., *Electrophoresis* 1997, 18, 360-368.
- [45] Figeys, D., Ning, Y., Aebersold, R., *Anal. Chem.* 1997, 69, 3153-3160.
- [46] Garrels, J. I., McLaughlin, C. S., Warner, J. R., Fletcher, B., Latter, G. I., Kobayashi, R., Schwender, B., Volpe, T., Anderson, D. S., Mesquita-Fuentes, R., Payne, W. E., *Electrophoresis* 1997, 18, 1347-1360.
- [47] Schuler, G. D., Boguski, M. S., Stewart, E. A., Stein, L. D., Gyapay, G., Rice, K., White, R. E., Rodriguez-Tome, P., Aggarwal, A., Bajorek, E., Bentolila, S., Birren, B. B., Butler, A., Castle, A. B., Chiannikulchai, N., Chu, A., Clee, C., Cowles, S., Day, P. J., Dibling, T., Drouot, N., Dunham, I., Duprat, S., Edwards, C., Fan, J.-B., Fang, N., Fizames, C., Garrett, C., Green, L., Hadley, D., Harris, M., Harrison, P., Brady, S., Hicks, A., Holloway, E., Hui, L., Hussain, S., Louis-Dit-Sully, C., Ma, J., MacGilvery, A., Mader, C., Maratukulam, A., Matise, T. C., McKusick, K. B., Morissette, J., Mungall, A., Muselet, D., Nusbaum, H. C., Page, D. C., Peck, A., Perkins, S., Piercy, M., Qin, F., Quackenbush, J., Ranby, S., Reif, T., Rozen, S., Sanders, X., She, X., Silva, J., Slonim, D. K., Soderlund, C., Sun, W.-L., Tabar, P., Thangarajah, T., Vega-Czarny, N., Vollrath, D., Voyticky, S., Wilmer, T., Wu, X., Adams, M. D., Auffray, C., Walter, N. A. R., Brandon, R., Dehejia, A., Goodfellow, P. N., Houlgatte, R., Hudson, J. R., Jr., Ide, S. E., Iorio, K. R., Lee, W. Y., Seki, N., Nagase, T., Ishikawa, K., Nomura, N., Phillips, C., Polymeropoulos, M. H., Sandusky, M., Schmitt, K., Berry, R., Swanson, K., Torres, R., Venter, J. C., Sikela, J. M., Beckmann, J. S., Weissenbach, J., Myers, R. M., Cox, D. R., James, M. R., Bentley, D., et al. *Science* 1996, 274, 540-546.
- [48] Sanchez, J.-C., Rouge, V., Pisteur, M., Ravier, F., Tonella, L., Moosmayer, M., Wilkins, M. R., Hochstrasser, D. F., *Electrophoresis* 1997, 18, 324-327.
- [49] Miller, I., Haynes, P., Gemeiner, M., Aebersold, R., Manzoni, C., Lovati, M. R., Vignati, M., Eberini, I., Gianazza, E., *Electrophoresis* 1998, 19, 1493-1500.
- [50] Garrels, J. I., *Nucleic Acids Res.* 1996, 24, 46-49.

# Impact of DNA Amplification on Gene Expression Patterns in Breast Cancer<sup>1,2</sup>

Elizabeth Hyman,<sup>3</sup> Päivikki Kauraniemi,<sup>3</sup> Sampsa Hautaniemi, Maija Wolf, Spyro Mousses, Ester Rozenblum, Markus Ringnér, Guido Sauter, Outi Monni, Abdel Elkahloun, Olli-P. Kallioniemi, and Anne Kallioniemi<sup>4</sup>

Howard Hughes Medical Institute-NIH Research Scholar, Bethesda, Maryland 20892 [E.H.]; Cancer Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland 20892 [E.H., P.K., S.H., M.W., S.M., E.R., M.R., A.E., O.K., A.K.]; Laboratory of Cancer Genetics, Institute of Medical Technology, University of Tampere and Tampere University Hospital, FIN-33520 Tampere, Finland [P.K., A.K.]; Signal Processing Laboratory, Tampere University of Technology, FIN-33101 Tampere, Finland [S.H.]; Institute of Pathology, University of Basel, CH-4003 Basel, Switzerland [G.S.]; and Biomedicum Biochip Center, Helsinki University Hospital, Biomedicum Helsinki, FIN-00014 Helsinki, Finland [O.M.]

## ABSTRACT

Genetic changes underlie tumor progression and may lead to cancer-specific expression of critical genes. Over 1100 publications have described the use of comparative genomic hybridization (CGH) to analyze the pattern of copy number alterations in cancer, but very few of the genes affected are known. Here, we performed high-resolution CGH analysis on cDNA microarrays in breast cancer and directly compared copy number and mRNA expression levels of 13,824 genes to quantitate the impact of genomic changes on gene expression. We identified and mapped the boundaries of 24 independent amplicons, ranging in size from 0.2 to 12 Mb. Throughout the genome, both high- and low-level copy number changes had a substantial impact on gene expression, with 44% of the highly amplified genes showing overexpression and 10.5% of the highly overexpressed genes being amplified. Statistical analysis with random permutation tests identified 270 genes whose expression levels across 14 samples were systematically attributable to gene amplification. These included most previously described amplified genes in breast cancer and many novel targets for genomic alterations, including the *HOXB7* gene, the presence of which in a novel amplicon at 17q21.3 was validated in 10.2% of primary breast cancers and associated with poor patient prognosis. In conclusion, CGH on cDNA microarrays revealed hundreds of novel genes whose overexpression is attributable to gene amplification. These genes may provide insights to the clonal evolution and progression of breast cancer and highlight promising therapeutic targets.

## INTRODUCTION

Gene expression patterns revealed by cDNA microarrays have facilitated classification of cancers into biologically distinct categories, some of which may explain the clinical behavior of the tumors (1-6). Despite this progress in diagnostic classification, the molecular mechanisms underlying gene expression patterns in cancer have remained elusive, and the utility of gene expression profiling in the identification of specific therapeutic targets remains limited.

Accumulation of genetic defects is thought to underlie the clonal evolution of cancer. Identification of the genes that mediate the effects of genetic changes may be important by highlighting transcripts that are actively involved in tumor progression. Such transcripts and their encoded proteins would be ideal targets for anticancer therapies, as demonstrated by the clinical success of new therapies against amplified oncogenes, such as *ERBB2* and *EGFR* (7, 8), in breast cancer and other solid tumors. Besides amplifications of known oncogenes, over

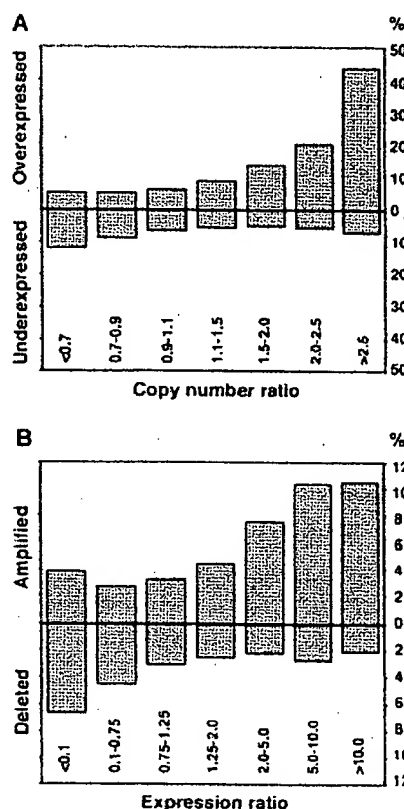


Fig. 1. Impact of gene copy number on global gene expression levels. A, percentage of over- and underexpressed genes (Y axis) according to copy number ratios (X axis). Threshold values used for over- and underexpression were  $>2.184$  (global upper 7% of the cDNA ratios) and  $<0.4826$  (global lower 7% of the expression ratios). B, percentage of amplified and deleted genes according to expression ratios. Threshold values for amplification and deletion were  $>1.5$  and  $<0.7$ .

20 recurrent regions of DNA amplification have been mapped in breast cancer by CGH<sup>3</sup> (9, 10). However, these amplicons are often large and poorly defined, and their impact on gene expression remains unknown.

We hypothesized that genome-wide identification of those gene expression changes that are attributable to underlying gene copy number alterations would highlight transcripts that are actively involved in the causation or maintenance of the malignant phenotype. To identify such transcripts, we applied a combination of cDNA and CGH microarrays to: (a) determine the global impact that gene copy number variation plays in breast cancer development and progression; and (b) identify and characterize those genes whose mRNA expres-

Received 5/29/02; accepted 8/28/02.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

<sup>1</sup> Supported in part by the Academy of Finland, Emil Aaltonen Foundation, the Finnish Cancer Society, the Pirkanmaa Cancer Society, the Pirkanmaa Cultural Foundation, the Finnish Breast Cancer Group, the Foundation for the Development of Laboratory Medicine, the Medical Research Fund of the Tampere University Hospital, the Foundation for Commercial and Technical Sciences, and the Swedish Research Council.

<sup>2</sup> Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org>).

<sup>3</sup> Contributed equally to this work.

<sup>4</sup> To whom requests for reprints should be addressed, at Laboratory of Cancer Genetics, Institute of Medical Technology, Lenkitie 6, FIN-33520 Tampere, Finland. Phone: 358-3247-4125; Fax: 358-3247-4168; E-mail: anne.kallioniemi@uta.fi.

<sup>5</sup> The abbreviations used are: CGH, comparative genomic hybridization; FISH, fluorescence in situ hybridization; RT-PCR, reverse transcription-PCR.

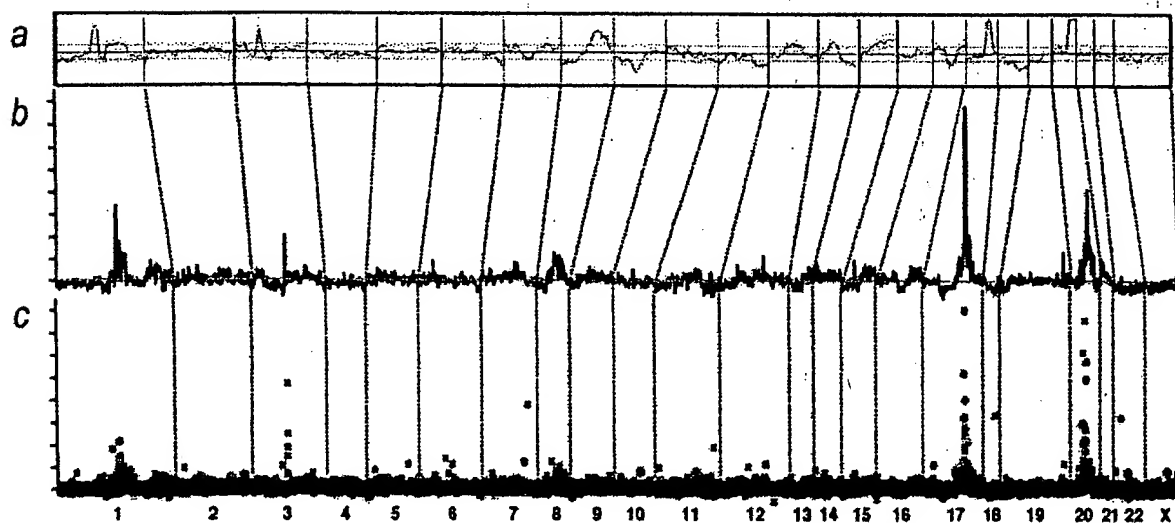


Fig. 2. Genome-wide copy number and expression analysis in the MCF-7 breast cancer cell line. *A*, chromosomal CGH analysis of MCF-7. The copy number ratio profile (blue line) across the entire genome from 1p telomere to Xq telomere is shown along with  $\pm 1$  SD (orange lines). The black horizontal line indicates a ratio of 1.0; red line, a ratio of 0.8; and green line, a ratio of 1.2. *B-C*, genome-wide copy number analysis in MCF-7 by CGH on cDNA microarray. The copy number ratios were plotted as a function of the position of the cDNA clones along the human genome. In *B*, individual data points are connected with a line, and a moving median of 10 adjacent clones is shown. Red horizontal line, the copy number ratio of 1.0. In *C*, individual data points are labeled by color coding according to cDNA expression ratios. The bright red dots indicate the upper 2%, the next 5% of the expression ratios in MCF-7 cells (overexpressed genes); bright green dots indicate the lowest 2%, and dark green dots, the next 5% of the expression ratios (underexpressed genes); the rest of the observations are shown with black crosses. The chromosome numbers are shown at the bottom of the figure, and chromosome boundaries are indicated with a dashed line.

sion is most significantly associated with amplification of the corresponding genomic template.

## MATERIALS AND METHODS

**Breast Cancer Cell Lines.** Fourteen breast cancer cell lines (BT-20, BT-474, HCC1428, Hs578t, MCF7, MDA-361, MDA-436, MDA-453, MDA-468, SKBR-3, T-47D, UACC812, ZR-75-1, and ZR-75-30) were obtained from the American Type Culture Collection (Manassas, VA). Cells were grown under recommended culture conditions. Genomic DNA and mRNA were isolated using standard protocols.

**Copy Number and Expression Analyses by cDNA Microarrays.** The preparation and printing of the 13,824 cDNA clones on glass slides were performed as described (11-13). Of these clones, 244 represented uncharacterized expressed sequence tags, and the remainder corresponded to known genes. CGH experiments on cDNA microarrays were done as described (14, 15). Briefly, 20  $\mu$ g of genomic DNA from breast cancer cell lines and normal human WBCs were digested for 14-18 h with *AluI* and *RsaI* (Life Technologies, Inc., Rockville, MD) and purified by phenol/chloroform extraction. Six  $\mu$ g of digested cell line DNAs were labeled with Cy3-dUTP (Amersham Pharmacia) and normal DNA with Cy5-dUTP (Amersham Pharmacia) using the Bioprime Labeling kit (Life Technologies, Inc.). Hybridization (14, 15) and posthybridization washes (13) were done as described. For the expression analyses, a standard reference (Universal Human Reference RNA; Stratagene, La Jolla, CA) was used in all experiments. Forty  $\mu$ g of reference RNA were labeled with Cy3-dUTP and 3.5  $\mu$ g of test mRNA with Cy5-dUTP, and the labeled cDNAs were hybridized on microarrays as described (13, 15). For both microarray analyses, a laser confocal scanner (Agilent Technologies, Palo Alto, CA) was used to measure the fluorescence intensities at the target locations using the DEARRAY software (16). After background subtraction, average intensities at each clone in the test hybridization were divided by the average intensity of the corresponding clone in the control hybridization. For the copy number analysis, the ratios were normalized on the basis of the distribution of ratios of all targets on the array and for the expression analysis on the basis of 88 housekeeping genes, which were spotted four times onto the array. Low quality measurements (*i.e.*, copy number data with mean reference intensity <100 fluorescent units, and expression data with both test and reference intensity <100 fluorescent units and/or with spot size <50 units)

were excluded from the analysis and were treated as missing values. The distributions of fluorescence ratios were used to define cutpoints for increased/decreased copy number. Genes with CGH ratio >1.43 (representing the upper 5% of the CGH ratios across all experiments) were considered to be amplified, and genes with ratio <0.73 (representing the lower 5%) were considered to be deleted.

**Statistical Analysis of CGH and cDNA Microarray Data.** To evaluate the influence of copy number alterations on gene expression, we applied the following statistical approach. CGH and cDNA calibrated intensity ratios were log-transformed and normalized using median centering of the values in each cell line. Furthermore, cDNA ratios for each gene across all 14 cell lines were median centered. For each gene, the CGH data were represented by a vector that was labeled 1 for amplification (ratio, >1.43) and 0 for no amplification. Amplification was correlated with gene expression using the signal-to-noise statistics (1). We calculated a weight,  $w_g$ , for each gene as follows:

$$w_g = \frac{m_{g1} - m_{g0}}{\sigma_{g1} + \sigma_{g0}}$$

where  $m_{g1}$ ,  $\sigma_{g1}$  and  $m_{g0}$ ,  $\sigma_{g0}$  denote the means and SDs for the expression levels for amplified and nonamplified cell lines, respectively. To assess the statistical significance of each weight, we performed 10,000 random permutations of the label vector. The probability that a gene had a larger or equal weight by random permutation than the original weight was denoted by  $\alpha$ . A low  $\alpha$  (<0.05) indicates a strong association between gene expression and amplification.

**Genomic Localization of cDNA Clones and Amplicon Mapping.** Each cDNA clone on the microarray was assigned to a Unigene cluster using the Unigene Build 141.<sup>6</sup> A database of genomic sequence alignment information for mRNA sequences was created from the August 2001 freeze of the University of California Santa Cruz's GoldenPath database.<sup>7</sup> The chromosome and bp positions for each cDNA clone were then retrieved by relating these data sets. Amplicons were defined as a CGH copy number ratio >2.0 in at least two adjacent clones in two or more cell lines or a CGH ratio >2.0 in at least three adjacent clones in a single cell line. The amplicon start and end positions were

<sup>6</sup> Internet address: [http://research.nhgri.nih.gov/microarray/downloadable\\_cdna.html](http://research.nhgri.nih.gov/microarray/downloadable_cdna.html).

<sup>7</sup> Internet address: [www.genome.ucsc.edu](http://www.genome.ucsc.edu).

Table 1 Summary of independent amplicons in 14 breast cancer cell lines by CGH microarray

Location	Start (Mb)	End (Mb)	Size (Mb)
1p13	132.79	132.94	0.2
1q21	173.92	177.25	3.3
1q22	179.28	179.57	0.3
3p14	71.94	74.66	2.7
7p12.1-7p11.2	55.62	60.95	5.3
7q31	125.73	130.96	5.2
7q32	140.01	140.68	0.7
8q21.11-8q21.13	86.45	92.46	6.0
8q21.3	98.45	103.05	4.6
8q23.3-8q24.14	129.88	142.15	12.3
8q24.22	151.21	152.16	1.0
9p13	38.65	39.25	0.6
13q22-q31	77.15	81.38	4.2
16q22	86.70	87.62	0.9
17q11	29.30	30.85	1.6
17q12-q21.2	39.79	42.80	3.0
17q21.32-q21.33	52.47	55.80	3.3
17q22-q23.3	63.81	69.70	5.9
17q23.3-q24.3	69.93	74.99	5.1
19q13	40.63	41.40	0.8
20q11.22	34.59	35.85	1.3
20q13.12	44.00	45.62	1.6
20q13.12-q13.13	46.45	49.43	3.0
20q13.2-q13.32	51.32	59.12	7.8

extended to include neighboring nonamplified clones (ratio, <1.5). The amplicon size determination was partially dependent on local clone density.

**FISH.** Dual-color interphase FISH to breast cancer cell lines was done as described (17). Bacterial artificial chromosome clone RP11-361K8 was labeled with SpectrumOrange (Vysis, Downers Grove, IL), and SpectrumOrange-labeled probe for *EGFR* was obtained from Vysis. SpectrumGreen-labeled chromosome 7 and 17 centromere probes (Vysis) were used as a reference. A tissue microarray containing 612 formalin-fixed, paraffin-embedded primary breast cancers (17) was applied in FISH analyses as described (18). The use of these specimens was approved by the Ethics Committee of the University of Basel and by the NIH. Specimens containing a 2-fold or higher increase in the number of test probe signals, as compared with corresponding centromere signals, in at least 10% of the tumor cells were considered to be amplified. Survival analysis was performed using the Kaplan-Meier method and the log-rank test.

**RT-PCR.** The *HOXB7* expression level was determined relative to *GAPDH*. Reverse transcription and PCR amplification were performed using Access RT-PCR System (Promega Corp., Madison, WI) with 10 ng of mRNA as a template. *HOXB7* primers were 5'-GAGCAGAGGGACTCGGACTT-3' and 5'-GCGTCAGGTAGCGATTGTAG-3'.

## RESULTS

**Global Effect of Copy Number on Gene Expression.** 13,824 arrayed cDNA clones were applied for analysis of gene expression and gene copy number (CGH microarrays) in 14 breast cancer cell lines. The results illustrate a considerable influence of copy number on gene expression patterns. Up to 44% of the highly amplified transcripts (CGH ratio, >2.5) were overexpressed (i.e., belonged to the global upper 7% of expression ratios), compared with only 6% for genes with normal copy number levels (Fig. 1A). Conversely, 10.5% of the transcripts with high-level expression (cDNA ratio, >10) showed increased copy number (Fig. 1B). Low-level copy number increases and decreases were also associated with similar, although less dramatic, outcomes on gene expression (Fig. 1).

**Identification of Distinct Breast Cancer Amplicons.** Base-pair locations obtained for 11,994 cDNAs (86.8%) were used to plot copy number changes as a function of genomic position (Fig. 2, Supplement Fig. A). The average spacing of clones throughout the genome was 267 kb. This high-resolution mapping identified 24 independent breast cancer amplicons, spanning from 0.2 to 12 Mb of DNA (Table 1). Several amplification sites detected previously by chromosomal

CGH were validated, with 1q21, 17q12-q21.2, 17q22-q23, 20q13.1, and 20q13.2 regions being most commonly amplified. Furthermore, the boundaries of these amplicons were precisely delineated. In addition, novel amplicons were identified at 9p13 (38.65-39.25 Mb), and 17q21.3 (52.47-55.80 Mb).

**Direct Identification of Putative Amplification Target Genes.** The cDNA/CGH microarray technique enables the direct correlation of copy number and expression data on a gene-by-gene basis throughout the genome. We directly annotated high-resolution CGH plots with gene expression data using color coding. Fig. 2C shows that most of the amplified genes in the MCF-7 breast cancer cell line at 1p13, 17q22-q23, and 20q13 were highly overexpressed. A view of chromosome 7 in the MDA-468 cell line implicates *EGFR* as the most highly overexpressed and amplified gene at 7p11-p12 (Fig. 3A). In BT-474, the two known amplicons at 17q12 and 17q22-q23 contained numerous highly overexpressed genes (Fig. 3B). In addition, several genes, including the homeobox genes *HOXB2* and *HOXB7*, were highly amplified in a previously undescribed independent amplicon at 17q21.3. *HOXB7* was systematically amplified (as validated by FISH, Fig. 3B, inset) as well as overexpressed (as verified by RT-PCR, data not shown) in BT-474, UACC812, and ZR-75-30 cells. Furthermore, this novel

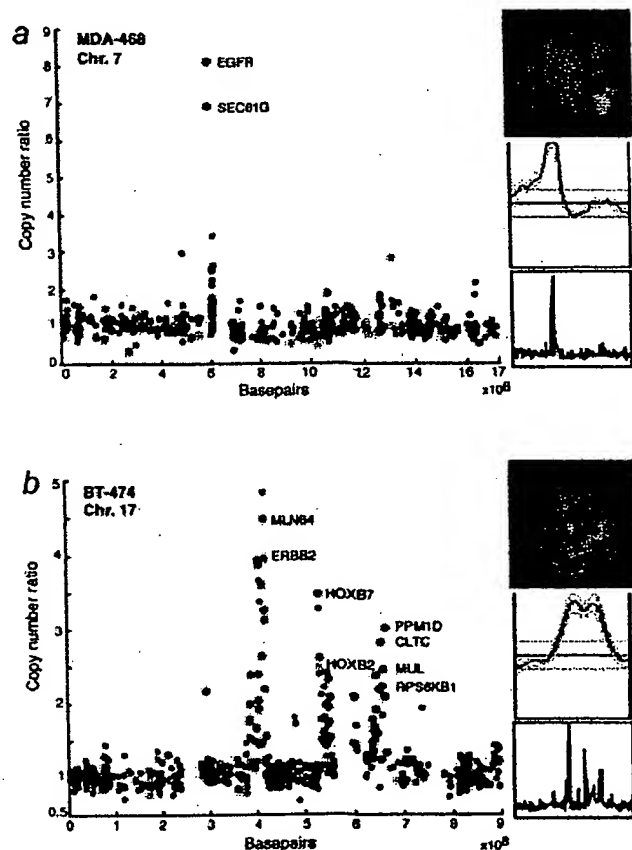
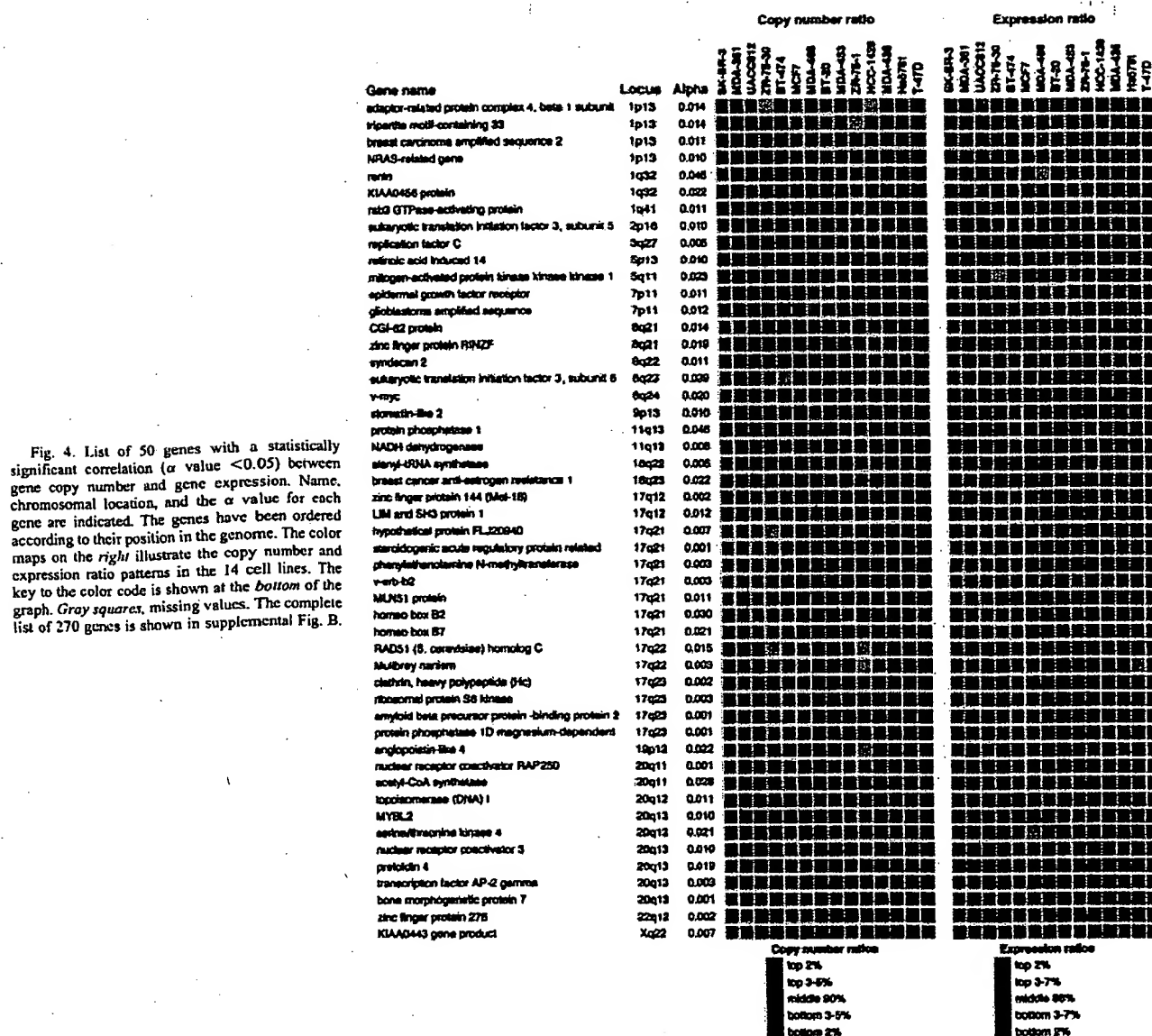


Fig. 3. Annotation of gene expression data on CGH microarray profiles. A, genes in the 7p11-p12 amplicon in the MDA-468 cell line are highly expressed (red dots) and include the *EGFR* oncogene. B, several genes in the 17q12, 17q21.3, and 17q23 amplicons in the BT-474 breast cancer cell line are highly overexpressed (red) and include the *HOXB7* gene. The data labels and color coding are as indicated for Fig. 2C. Insets show chromosomal CGH profiles for the corresponding chromosomes and validation of the increased copy number by interphase FISH using *EGFR* (red) and chromosome 7 centromere probe (green) to MDA-468 (A) and *HOXB7*-specific probe (red) and chromosome 17 centromere (green) to BT-474 cells (B).



# GENE EXPRESSION PATTERNS IN BREAST CANCER



amplification was validated to be present in 10.2% of 363 primary breast cancers by FISH to a tissue microarray and was associated with poor prognosis of the patients ( $P = 0.001$ ).

**Statistical Identification and Characterization of 270 Highly Expressed Genes in Amplicons.** Statistical comparison of expression levels of all genes as a function of gene amplification identified 270 genes whose expression was significantly influenced by copy number across all 14 cell lines (Fig. 4, Supplemental Fig. B). According to the gene ontology data,<sup>8</sup> 91 of the 270 genes represented hypothetical proteins or genes with no functional annotation, whereas 179 had associated functional information available. Of these, 151 (84%) are implicated in apoptosis, cell proliferation, signal transduction, and transcription, whereas 28 (16%) had functional annotations that could not be directly linked with cancer.

## DISCUSSION

The importance of recurrent gene and chromosome copy number changes in the development and progression of solid tumors has been characterized in  $>1000$  publications applying CGH<sup>9</sup> (9, 10), as well as in a large number of other molecular cytogenetic, cytogenetic, and molecular genetic studies. The effects of these somatic genetic changes on gene expression levels have remained largely unknown, although a few studies have explored gene expression changes occurring in specific amplicons (15, 19–21). Here, we applied genome-wide cDNA microarrays to identify transcripts whose expression changes were attributable to underlying gene copy number alterations in breast cancer.

The overall impact of copy number on gene expression patterns was substantial with the most dramatic effects seen in the case of high-

<sup>8</sup> Internet address: <http://www.geneontology.org/>.

<sup>9</sup> Internet address: <http://www.ncbi.nlm.nih.gov/entrez>.



level copy number increase. Low-level copy number gains and losses also had a significant influence on expression levels of genes in the regions affected, but these effects were more subtle on a gene-by-gene basis than those of high-level amplifications. However, the impact of low-level gains on the dysregulation of gene expression patterns in cancer may be equally important if not more important than that of high-level amplifications. Aneuploidy and low-level gains and losses of chromosomal arms represent the most common types of genetic alterations in breast and other cancers and, therefore, have an influence on many genes. Our results in breast cancer extend the recent studies on the impact of aneuploidy on global gene expression patterns in yeast cells, acute myeloid leukemia, and a prostate cancer model system (22–24).

The CGH microarray analysis identified 24 independent breast cancer amplicons. We defined the precise boundaries for many amplicons detected previously by chromosomal CGH (9, 10, 25, 26) and also discovered novel amplicons that had not been detected previously, presumably because of their small size (only 1–2 Mb) or close proximity to other larger amplicons. One of these novel amplicons involved the homeobox gene region at 17q21.3 and led to the overexpression of the *HOXB7* and *HOXB2* genes. The homeodomain transcription factors are known to be key regulators of embryonic development and have been occasionally reported to undergo aberrant expression in cancer (27, 28). *HOXB7* transfection induced cell proliferation in melanoma, breast, and ovarian cancer cells and increased tumorigenicity and angiogenesis in breast cancer (29–32). The present results imply that gene amplification may be a prominent mechanism for overexpressing *HOXB7* in breast cancer and suggest that *HOXB7* contributes to tumor progression and confers an aggressive disease phenotype in breast cancer. This view is supported by our finding of amplification of *HOXB7* in 10% of 363 primary breast cancers, as well as an association of amplification with poor prognosis of the patients.

We carried out a systematic search to identify genes whose expression levels across all 14 cell lines were attributable to amplification status. Statistical analysis revealed 270 such genes (representing ~2% of all genes on the array), including not only previously described amplified genes, such as *HER-2*, *MYC*, *EGFR*, ribosomal protein S6 kinase, and *AIB3*, but also numerous novel genes such as *NRAS-related gene* (1p13), *syndecan-2* (8q22), and *bone morphogenic protein* (20q13.1), whose activation by amplification may similarly promote breast cancer progression. Most of the 270 genes have not been implicated previously in breast cancer development and suggest novel pathogenetic mechanisms. Although we would not expect all of them to be causally involved, it is intriguing that 84% of the genes with associated functional information were implicated in apoptosis, cell proliferation, signal transduction, transcription, or other cellular processes that could directly imply a possible role in cancer progression. Therefore, a detailed characterization of these genes may provide biological insights to breast cancer progression and might lead to the development of novel therapeutic strategies.

In summary, we demonstrate application of cDNA microarrays to the analysis of both copy number and expression levels of over 12,000 transcripts throughout the breast cancer genome, roughly once every 267 kb. This analysis provided: (a) evidence of a prominent global influence of copy number changes on gene expression levels; (b) a high-resolution map of 24 independent amplicons in breast cancer; and (c) identification of a set of 270 genes, the overexpression of which was statistically attributable to gene amplification. Characterization of a novel amplicon at 17q21.3 implicated amplification and overexpression of the *HOXB7* gene in breast cancer, including a clinical association

between *HOXB7* amplification and poor patient prognosis. Overall, our results illustrate how the identification of genes activated by gene amplification provides a powerful approach to highlight genes with an important role in cancer as well as to prioritize and validate putative targets for therapy development.

## REFERENCES

- Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., and Lander, E. S. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* (Wash. DC), 286: 531–537, 1999.
- Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* (Lond.), 403: 503–511, 2000.
- Bittner, M., Meltzer, P., Chen, Y., Jiang, Y., Seftor, E., Hendrix, M., Radmacher, M., Simon, R., Yakhini, Z., Ben-Dor, A., et al. Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature* (Lond.), 406: 536–540, 2000.
- Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., et al. Molecular portraits of human breast tumours. *Nature* (Lond.), 406: 747–752, 2000.
- Dhanasekaran, S. M., Barrette, T. R., Ghosh, D., Shah, R., Varambally, S., Kumar, K., Pienta, K. J., Rubin, M. A., and Chinnaiyan, A. M. Delineation of prognostic biomarkers in prostate cancer. *Nature* (Lond.), 412: 822–826, 2001.
- Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. USA*, 98: 10869–10874, 2001.
- Ross, J. S., and Fletcher, J. A. The *HER-2/neu* oncogene: prognostic factor, predictive factor and target for therapy. *Semin. Cancer Biol.*, 9: 125–138, 1999.
- Arteaga, C. L. The epidermal growth factor receptor: from mutant oncogene in nonhuman cancers to therapeutic target in human neoplasia. *J. Clin. Oncol.*, 19: 32–40, 2001.
- Knuutila, S., Björkqvist, A. M., Autio, K., Tarkkanen, M., Wolf, M., Monni, O., Szymanska, J., Larramendy, M. L., Tapper, J., Pere, H., El-Rifai, W., et al. DNA copy number amplifications in human neoplasms: review of comparative genomic hybridization studies. *Am. J. Pathol.*, 152: 1107–1123, 1998.
- Knuutila, S., Autio, K., and Aalto, Y. Online access to CGH data of DNA sequence copy number changes. *Am. J. Pathol.*, 157: 689, 2000.
- DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A., and Trent, J. M. Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat. Genet.*, 14: 457–460, 1996.
- Shalon, D., Smith, S. J., and Brown, P. O. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res.*, 6: 639–645, 1996.
- Mousses, S., Bittner, M. L., Chen, Y., Dougherty, E. R., Baxevanis, A., Meltzer, P. S., and Trent, J. M. Gene expression analysis by cDNA microarrays. In: F. J. Livesey and S. P. Hunt (eds.), *Functional Genomics*. pp. 113–137. Oxford: Oxford University Press, 2000.
- Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Bostein, D., and Brown, P. O. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.*, 23: 41–46, 1999.
- Monni, O., Bärnlund, M., Mousses, S., Kononen, J., Sauter, G., Heiskanen, M., Paavola, P., Avela, K., Chen, Y., Bittner, M. L., and Kallioniemi, A. Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. *Proc. Natl. Acad. Sci. USA*, 98: 5711–5716, 2001.
- Chen, Y., Dougherty, E. R., and Bittner, M. L. Ratio-based decisions and the quantitative analysis of cDNA microarray images. *J. Biomed. Optics*, 2: 364–374, 1997.
- Bärnlund, M., Forozan, F., Kononen, J., Bubendorf, L., Chen, Y., Bittner, M. L., Thorst, J., Haas, P., Bucher, C., Sauter, G., et al. Detecting activation of ribosomal protein S6 kinase by complementary DNA and tissue microarray analysis. *J. Natl. Cancer Inst.*, 92: 1252–1259, 2000.
- Andersen, C. L., Hostetter, G., Grigoryan, A., Sauter, G., and Kallioniemi, A. Improved procedure for fluorescence *in situ* hybridization on tissue microarrays. *Cytometry*, 45: 83–86, 2001.
- Kauraniemi, P., Bärnlund, M., Monni, O., and Kallioniemi, A. New amplified and highly expressed genes discovered in the ERBB2 amplicon in breast cancer by cDNA microarrays. *Cancer Res.*, 61: 8235–8240, 2001.
- Clark, J., Edwards, S., John, M., Flohr, P., Gordon, T., Maillard, K., Giddings, I., Brown, C., Bagherzadeh, A., Campbell, C., Shipley, J., Wooster, R., and Cooper, C. S. Identification of amplified and expressed genes in breast cancer by comparative hybridization onto microarrays of randomly selected cDNA clones. *Genes Chromosomes Cancer*, 34: 104–114, 2002.
- Varis, A., Wolf, M., Monni, O., Vakkari, M. L., Kokkola, A., Moskaluk, C., Frierson, H., Powell, S. M., Knuutila, S., Kallioniemi, A., and El-Rifai, W. Targets of gene amplification and overexpression at 17q in gastric cancer. *Cancer Res.*, 62: 2625–2629, 2002.
- Hughes, T. R., Roberts, C. J., Dai, H., Jones, A. R., Meyer, M. R., Slade, D., Burchard, J., Dow, S., Ward, T. R., Kidd, M. J., Friend, S. H., and Marton, M. J.

- Widespread aneuploidy revealed by DNA microarray expression profiling. *Nat. Genet.*, 25: 333-337, 2000.
23. Virtaneva, K., Wright, F. A., Tanner, S. M., Yuan, B., Lemon, W. J., Caligiuri, M. A., Bloomfield, C. D., de La Chapelle, A., and Krahe, R. Expression profiling reveals fundamental biological differences in acute myeloid leukemia with isolated trisomy 8 and normal cytogenetics. *Proc. Natl. Acad. Sci. USA*, 98: 1124-1129, 2001.
24. Phillips, J. L., Hayward, S. W., Wang, Y., Vasselli, J., Pavlovich, C., Padilla-Nash, H., Pezullo, J. R., Ghadimi, B. M., Grossfeld, G. D., Rivera, A., Linchan, W. M., Cunha, G. R., and Ried, T. The consequences of chromosomal aneuploidy on gene expression profiles in a cell line model for prostate carcinogenesis. *Cancer Res.*, 61: 8143-8149, 2001.
25. Bärklund, M., Tirkkonen, M., Forozan, F., Tanner, M. M., Kallioniemi, O. P., and Kallioniemi, A. Increased copy number at 17q22-q24 by CGH in breast cancer is due to high-level amplification of two separate regions. *Genes Chromosomes Cancer*, 20: 372-376, 1997.
26. Tanner, M. M., Tirkkonen, M., Kallioniemi, A., Isola, J., Kuukasjärvi, T., Collins, C., Kowbel, D., Guan, X. Y., Trent, J., Gray, J. W., Meltzer, P., and Kallioniemi, O. P. Independent amplification and frequent co-amplification of three nonsynthetic regions on the long arm of chromosome 20 in human breast cancer. *Cancer Res.*, 56: 3441-3445, 1996.
27. Cillo, C., Faiella, A., Cantile, M., and Boncinelli, E. Homeobox genes and cancer. *Exp. Cell Res.*, 248: 1-9, 1999.
28. Cillo, C., Cantile, M., Faiella, A., and Boncinelli, E. Homeobox genes in normal and malignant cells. *J. Cell. Physiol.*, 188: 161-169, 2001.
29. Care, A., Silvani, A., Meccia, E., Mattia, G., Stoppacciaro, A., Parmiani, G., Peschle, C., and Colombo, M. P. HOXB7 constitutively activates basic fibroblast growth factor in melanomas. *Mol. Cell. Biol.*, 16: 4842-4851, 1996.
30. Care, A., Silvani, A., Meccia, E., Mattia, G., Peschle, C., and Colombo, M. P. Transduction of the SkBr3 breast carcinoma cell line with the HOXB7 gene induces bFGF expression, increases cell proliferation and reduces growth factor dependence. *Oncogene*, 16: 3285-3289, 1998.
31. Care, A., Felicetti, F., Meccia, E., Bottero, L., Parenza, M., Stoppacciaro, A., Peschle, C., and Colombo, M. P. HOXB7: a key factor for tumor-associated angiogenic switch. *Cancer Res.*, 61: 6532-6539, 2001.
32. Naora, H., Yang, Y. Q., Montz, F. J., Seidman, J. D., Kurman, R. J., and Roden, R. B. A serologically identified tumor antigen encoded by a homeobox gene promotes growth of ovarian epithelial cells. *Proc. Natl. Acad. Sci. USA*, 98: 4060-4065, 2001.

# Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors

Jonathan R. Pollack<sup>\*†‡</sup>, Therese Sørlie<sup>§</sup>, Charles M. Perou<sup>¶</sup>, Christian A. Rees<sup>||\*</sup>, Stefanie S. Jeffrey<sup>††</sup>, Per E. Lønning<sup>‡‡</sup>, Robert Tibshirani<sup>§§</sup>, David Botstein<sup>||</sup>, Anne-Lise Børresen-Dale<sup>§</sup>, and Patrick O. Brown<sup>††¶</sup>

Departments of <sup>\*</sup>Pathology, <sup>§</sup>Genetics, <sup>††</sup>Surgery, <sup>§§</sup>Health Research and Policy, and <sup>¶¶</sup>Biochemistry, and <sup>||</sup>Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA 94305; <sup>§</sup>Department of Genetics, Norwegian Radium Hospital, Montebello, N-0310 Oslo, Norway; <sup>††</sup>Department of Medicine (Oncology), Haukeland University Hospital, N-5021 Bergen, Norway; and <sup>‡‡</sup>Department of Genetics and Ureberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC 27599

Contributed by Patrick O. Brown, August 6, 2002

Genomic DNA copy number alterations are key genetic events in the development and progression of human cancers. Here we report a genome-wide microarray comparative genomic hybridization (array CGH) analysis of DNA copy number variation in a series of primary human breast tumors. We have profiled DNA copy number alteration across 6,691 mapped human genes, in 44 predominantly advanced, primary breast tumors and 10 breast cancer cell lines. While the overall patterns of DNA amplification and deletion corroborate previous cytogenetic studies, the high-resolution (gene-by-gene) mapping of amplicon boundaries and the quantitative analysis of amplicon shape provide significant improvement in the localization of candidate oncogenes. Parallel microarray measurements of mRNA levels reveal the remarkable degree to which variation in gene copy number contributes to variation in gene expression in tumor cells. Specifically, we find that 62% of highly amplified genes show moderately or highly elevated expression, that DNA copy number influences gene expression across a wide range of DNA copy number alterations (deletion, low-, mid- and high-level amplification), that on average, a 2-fold change in DNA copy number is associated with a corresponding 1.5-fold change in mRNA levels, and that overall, at least 12% of all the variation in gene expression among the breast tumors is directly attributable to underlying variation in gene copy number. These findings provide evidence that widespread DNA copy number alteration can lead directly to global deregulation of gene expression, which may contribute to the development or progression of cancer.

Conventional cytogenetic techniques, including comparative genomic hybridization (CGH) (1), have led to the identification of a number of recurrent regions of DNA copy number alteration in breast cancer cell lines and tumors (2–4). While some of these regions contain known or candidate oncogenes [e.g., FGFR1 (8p11), MYC (8q24), CCND1 (11q13), ERBB2 (17q12), and ZNF217 (20q13)] and tumor suppressor genes [RB1 (13q14) and TP53 (17p13)], the relevant gene(s) within other regions (e.g., gain of 1q, 8q22, and 17q22–24, and loss of 8p) remain to be identified. A high-resolution genome-wide map, delineating the boundaries of DNA copy number alterations in tumors, should facilitate the localization and identification of oncogenes and tumor suppressor genes in breast cancer. In this study, we have created such a map, using array-based CGH (5–7) to profile DNA copy number alteration in a series of breast cancer cell lines and primary tumors.

An unresolved question is the extent to which the widespread DNA copy number changes that we and others have identified in breast tumors alter expression of genes within involved regions. Because we had measured mRNA levels in parallel in the same samples (8), using the same DNA microarrays, we had an opportunity to explore on a genomic scale the relationship between DNA copy number changes and gene expression. From

this analysis, we have identified a significant impact of widespread DNA copy number alteration on the transcriptional programs of breast tumors.

## Materials and Methods

**Tumors and Cell Lines.** Primary breast tumors were predominantly large (>3 cm), intermediate-grade, infiltrating ductal carcinomas, with more than 50% being lymph node positive. The fraction of tumor cells within specimens averaged at least 50%. Details of individual tumors have been published (8, 9), and are summarized in Table 1, which is published as supporting information on the PNAS web site, [www.pnas.org](http://www.pnas.org). Breast cancer cell lines were obtained from the American Type Culture Collection. Genomic DNA was isolated either using Qiagen genomic DNA columns, or by phenol/chloroform extraction followed by ethanol precipitation.

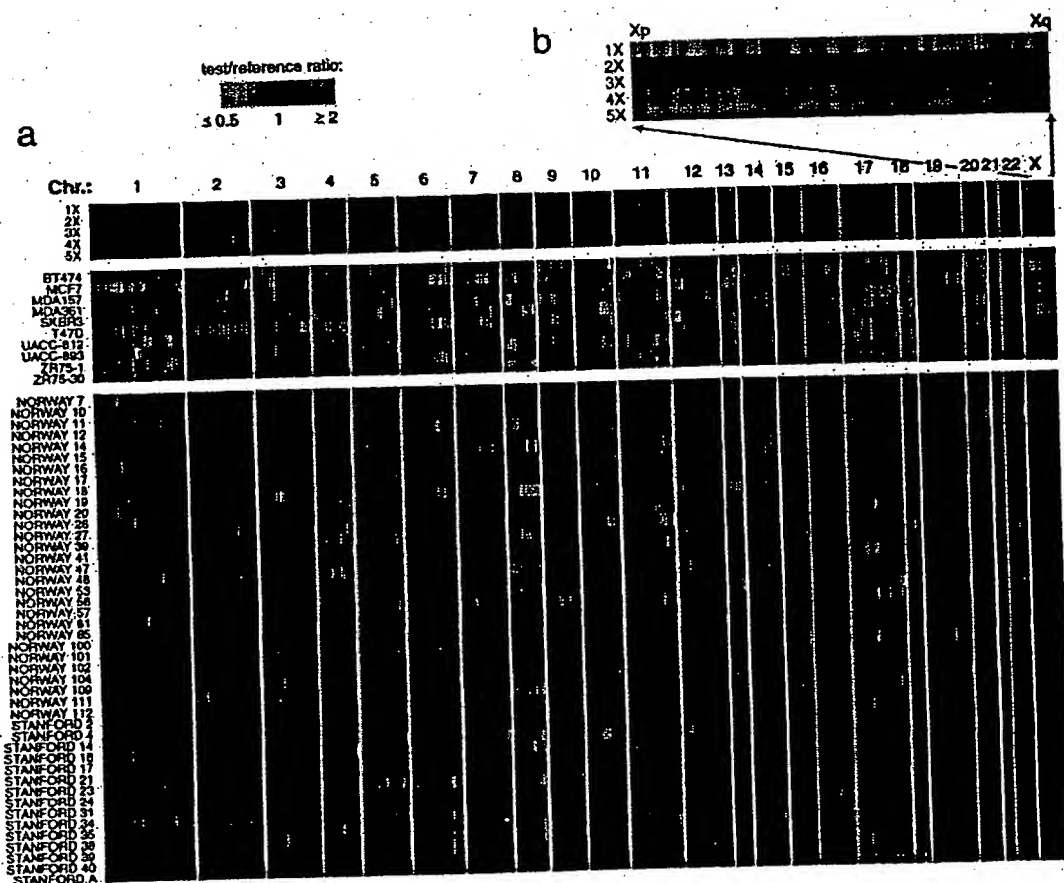
**DNA Labeling and Microarray Hybridizations.** Genomic DNA labeling and hybridizations were performed essentially as described in Pollack *et al.* (7), with slight modifications. Two micrograms of DNA was labeled in a total volume of 50 microliters and the volumes of all reagents were adjusted accordingly. “Test” DNA (from tumors and cell lines) was fluorescently labeled (Cy5) and hybridized to a human cDNA microarray containing 6,691 different mapped human genes (i.e., UniGene clusters). The “reference” (labeled with Cy3) for each hybridization was normal female leukocyte DNA from a single donor. The fabrication of cDNA microarrays and the labeling and hybridization of mRNA samples have been described (8).

**Data Analysis and Map Positions.** Hybridized arrays were scanned on a GenePix scanner (Axon Instruments, Foster City, CA), and fluorescence ratios (test/reference) calculated using SCANALYZE software (available at <http://rana.lbl.gov>). Fluorescence ratios were normalized for each array by setting the average log fluorescence ratio for all array elements equal to 0. Measurements with fluorescence intensities more than 20% above background were considered reliable. DNA copy number profiles that deviated significantly from background ratios measured in normal genomic DNA control hybridizations were interpreted as evidence of real DNA copy number alteration (see *Estimating Significance of Altered Fluorescence Ratios* in the supporting information). When indicated, DNA copy number profiles are displayed as a moving average (symmetric 5-nearest neighbors). Map positions for arrayed human cDNAs were assigned by

Abbreviation: CGH, comparative genomic hybridization.

<sup>†</sup>To whom reprint requests should be addressed at: Department of Pathology, Stanford University School of Medicine, CCSR Building, Room 3245A, 269 Campus Drive, Stanford, CA 94305-5176. E-mail: [pollack1@stanford.edu](mailto:pollack1@stanford.edu).

<sup>¶¶</sup>Present address: Zymyx Inc., Hayward, CA 94545.



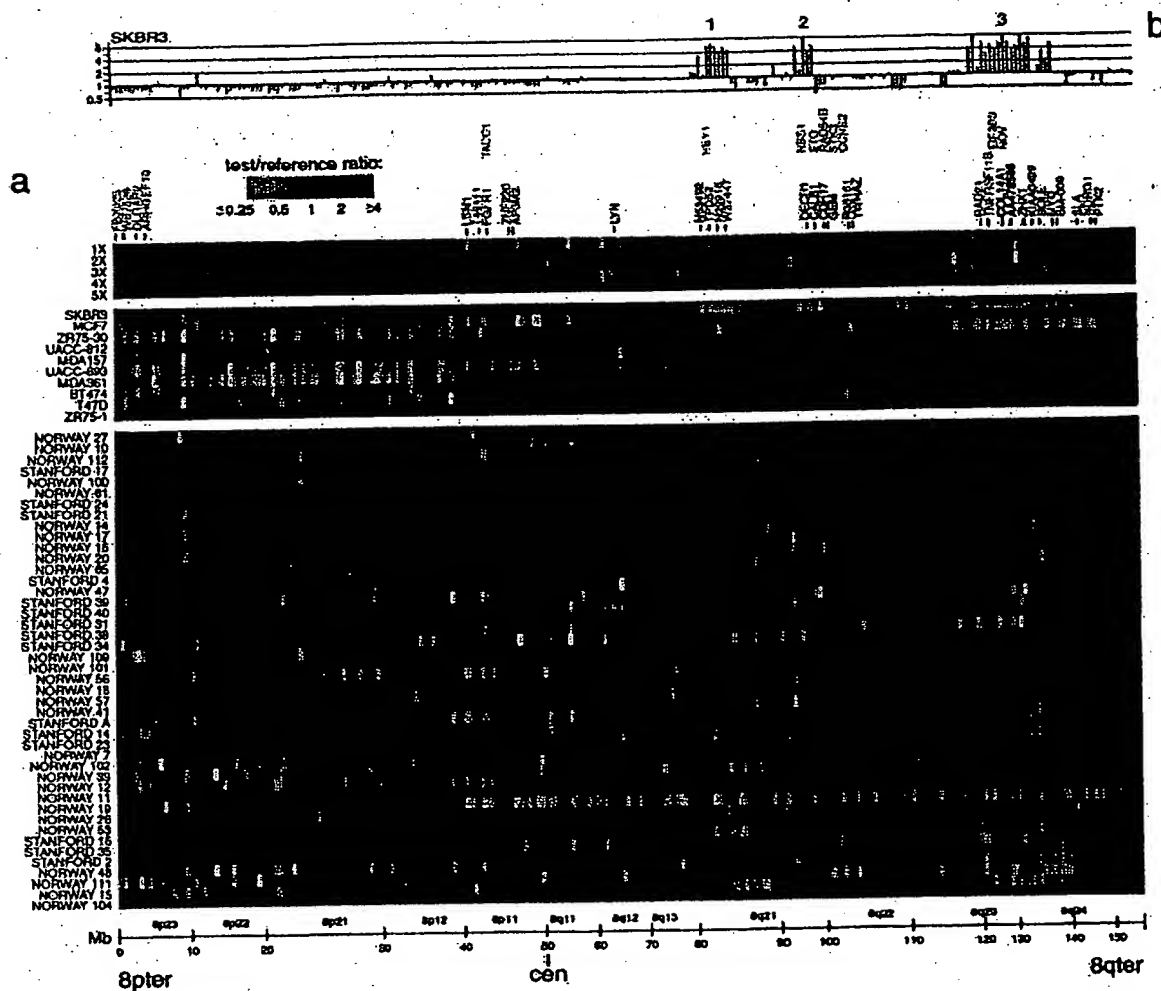
**Fig. 1.** Genome-wide measurement of DNA copy number alteration by array CGH. (a) DNA copy number profiles are illustrated for cell lines containing different numbers of X chromosomes, for breast cancer cell lines, and for breast tumors. Each row represents a different cell line or tumor, and each column represents one of 6,691 different mapped human genes present on the microarray, ordered by genome map position from 1pter through Xqter. Moving average (symmetric 5-nearest neighbors) fluorescence ratios (test/reference) are depicted using a log<sub>2</sub>-based pseudocolor scale (indicated), such that red luminescence reflects fold-amplification, green luminescence reflects fold-deletion, and black indicates no change (gray indicates poorly measured data). (b) Enlarged view of DNA copy number profiles across the X chromosome, shown for cell lines containing different numbers of X chromosomes.

identifying the starting position of the best and longest match of any DNA sequence represented in the corresponding UniGene cluster (10) against the "Golden Path" genome assembly (<http://genome.ucsc.edu/>; Oct 7, 2000 Freeze). For UniGene clusters represented by multiple arrayed elements, mean fluorescence ratios (for all elements representing the same UniGene cluster) are reported. For mRNA measurements, fluorescence ratios are "mean-centered" (i.e., reported relative to the mean ratio across the 44 tumor samples). The data set described here can be accessed in its entirety in the supporting information.

## Results

We performed CGH on 44 predominantly locally advanced, primary breast tumors and 10 breast cancer cell lines, using cDNA microarrays containing 6,691 different mapped human genes (Fig. 1a; also see *Materials and Methods* for details of microarray hybridizations). To take full advantage of the improved spatial resolution of array CGH, we ordered (fluorescence ratios for) the 6,691 cDNAs according to the "Golden Path" (<http://genome.ucsc.edu/>) genome assembly of the draft human genome sequences (11). In so doing, arrayed cDNAs not only themselves represent genes of potential interest (e.g., candidate oncogenes within amplicons), but also provide precise genetic landmarks for chromosomal regions of amplification and

deletion. Parallel analysis of DNA from cell lines containing different numbers of X chromosomes (Fig. 1b), as we did before (7), demonstrated the sensitivity of our method to detect single-copy loss (45, XO), and 1.5- (47,XXX), 2- (48,XXXX), or 2.5-fold (49,XXXXXX) gains (also see Fig. 5, which is published as supporting information on the PNAS web site). Fluorescence ratios were linearly proportional to copy number ratios, which were slightly underestimated, in agreement with previous observations (7). Numerous DNA copy number alterations were evident in both the breast cancer cell lines and primary tumors (Fig. 1a), detected in the tumors despite the presence of euploid non-tumor cell types; the magnitudes of the observed changes were generally lower in the tumor samples. DNA copy-number alterations were found in every cancer cell line and tumor, and on every human chromosome in at least one sample. Recurrent regions of DNA copy number gain and loss were readily identifiable. For example, gains within 1q, 8q, 17q, and 20q were observed in a high proportion of breast cancer cell lines/tumors (90%/69%, 100%/47%, 100%/60%, and 90%/44%, respectively), as were losses within 1p, 3p, 8p, and 13q (80%/24%, 80%/22%, 80%/22%, and 70%/18%, respectively), consistent with published cytogenetic studies (refs. 2-4; a complete listing of gains/losses is provided in Tables 2 and 3, which are published as supporting information on the PNAS web site). The total



**Fig. 2.** DNA copy number alteration across chromosome 8 by array CGH. (a) DNA copy number profiles are illustrated for cell lines containing different numbers of X chromosomes, for breast cancer cell lines, and for breast tumors. Breast cancer cell lines and tumors are separately ordered by hierarchical clustering to highlight recurrent copy number changes. The 241 genes present on the microarrays and mapping to chromosome 8 are ordered by position along the chromosome. Fluorescence ratios (test/reference) are depicted by a log<sub>2</sub> pseudocolor scale (indicated). Selected genes are indicated with color-coded text (red, increased; green, decreased; black, no change; gray, not well measured) to reflect correspondingly altered mRNA levels (observed in the majority of the subset of samples displaying the DNA copy number change). The map positions for genes of interest that are not represented on the microarray are indicated in the row above those genes represented on the array. (b) Graphical display of DNA copy number profile for breast cancer cell line SKBR3. Fluorescence ratios (tumor/normal) are plotted on a log<sub>2</sub> scale for chromosome 8 genes, ordered along the chromosome.

number of genomic alterations (gains and losses) was found to be significantly higher in breast tumors that were high grade ( $P = 0.008$ ), consistent with published CGH data (3), estrogen receptor negative ( $P = 0.04$ ), and harboring TP53 mutations ( $P = 0.0006$ ) (see Table 4; which is published as supporting information on the PNAS web site).

The improved spatial resolution of our array CGH analysis is illustrated for chromosome 8, which displayed extensive DNA copy number alteration in our series. A detailed view of the variation in the copy number of 241 genes mapping to chromosome 8 revealed multiple regions of recurrent amplification; each of these potentially harbors a different known or previously uncharacterized oncogene (Fig. 2a). The complexity of amplicon structure is most easily appreciated in the breast cancer cell line SKBR3. Although a conventional CGH analysis of 8q in SKBR3 identified only two distinct regions of amplification (12), we observed three distinct regions of high-level amplification (labeled 1–3 in Fig. 2b). For each of these regions we can define the

boundaries of the interval recurrently amplified in the tumors we examined; in each case, known or plausible candidate oncogenes can be identified (a description of these regions, as well as the recurrently amplified regions on chromosomes 17 and 20, can be found in Figs. 6 and 7, which are published as supporting information on the PNAS web site).

For a subset of breast cancer cell lines and tumors (4 and 37, respectively), and a subset of arrayed genes (6,095), mRNA levels were quantitatively measured in parallel by using cDNA microarrays (8). The parallel assessment of mRNA levels is useful in the interpretation of DNA copy number changes. For example, the highly amplified genes that are also highly expressed are the strongest candidate oncogenes within an amplicon. Perhaps more significantly, our parallel analysis of DNA copy number changes and mRNA levels provides us the opportunity to assess the global impact of widespread DNA copy number alteration on gene expression in tumor cells.

A strong influence of DNA copy number on gene expression is evident in an examination of the pseudocolor representations

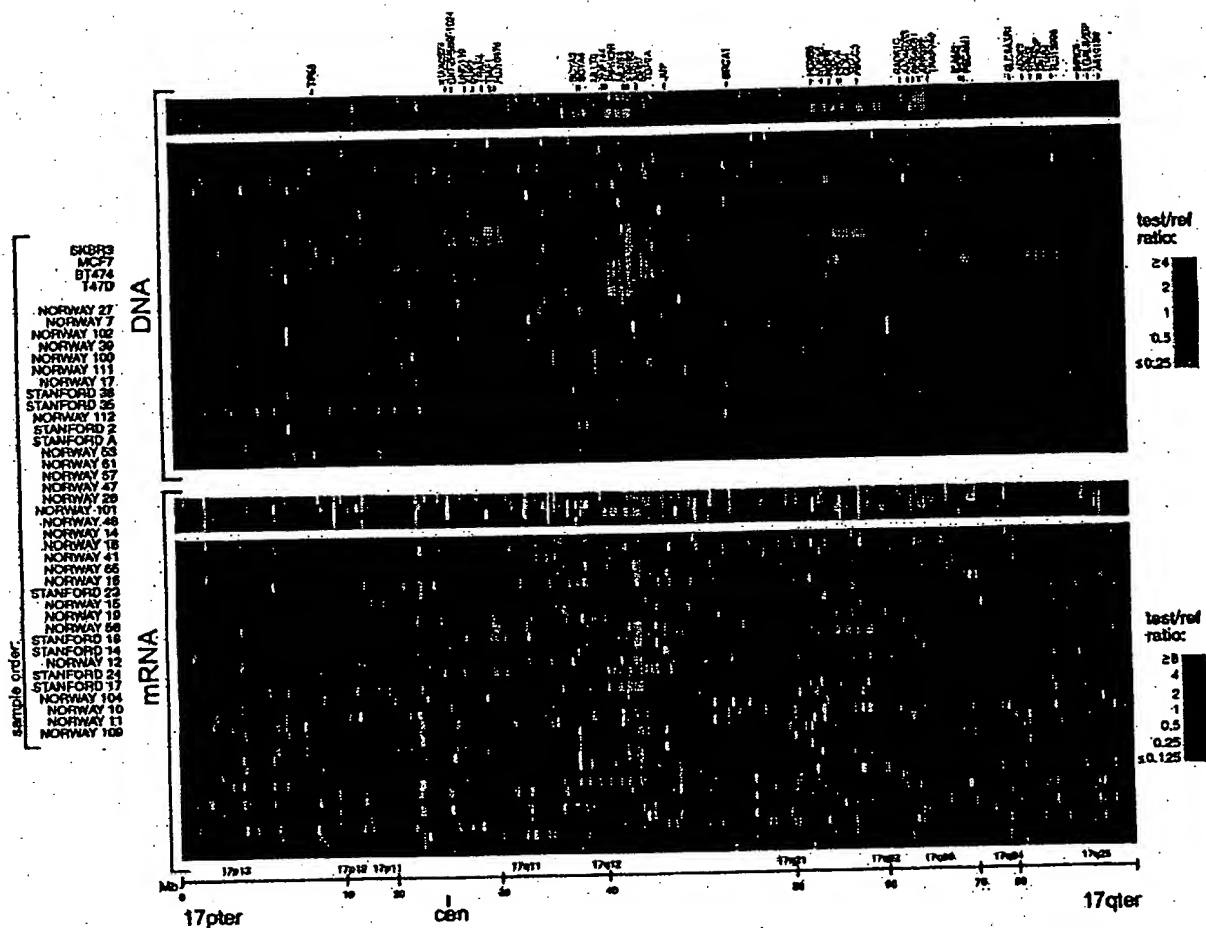
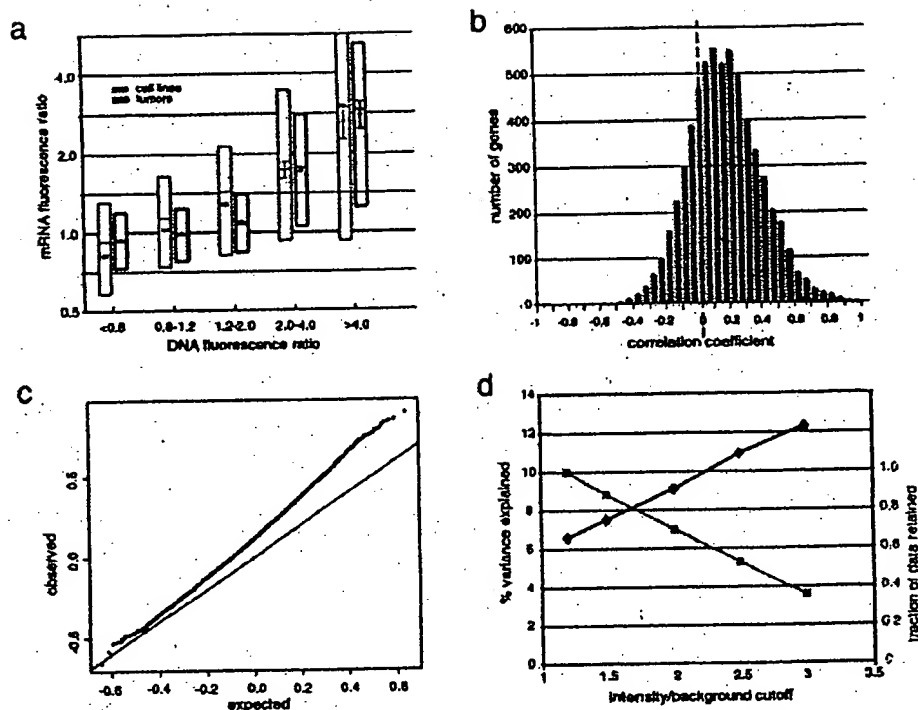


Fig. 3. Concordance between DNA copy number and gene expression across chromosome 17. DNA copy number alteration (Upper) and mRNA levels (Lower) are illustrated for breast cancer cell lines and tumors. Breast cancer cell lines and tumors are separately ordered by hierarchical clustering (Upper), and the identical sample order is maintained (Lower). The 354 genes present on the microarrays and mapping to chromosome 17, and for which both DNA copy number and mRNA levels were determined, are ordered by position along the chromosome; selected genes are indicated in color-coded text (see Fig. 2 legend). Fluorescence ratios (test/reference) are depicted by separate  $\log_2$  pseudocolor scales (indicated).

of DNA copy number and mRNA levels for genes on chromosome 17 (Fig. 3). The overall patterns of gene amplification and elevated gene expression are quite concordant; i.e., a significant fraction of highly amplified genes appear to be correspondingly highly expressed. The concordance between high-level amplification and increased gene expression is not restricted to chromosome 17. Genome-wide, of 117 high-level DNA amplifications (fluorescence ratios  $>4$ , and representing 91 different genes), 62% (representing 54 different genes; see Table 5, which is published as supporting information on the PNAS web site) are found associated with at least moderately elevated mRNA levels (mean-centered fluorescence ratios  $>2$ ), and 42% (representing 36 different genes) are found associated with comparably highly elevated mRNA levels (mean-centered fluorescence ratios  $>4$ ).

To determine the extent to which DNA deletion and lower-level amplification (in addition to high-level amplification) are also associated with corresponding alterations in mRNA levels, we performed three separate analyses on the complete data set (4 cell lines and 37 tumors, across 6,095 genes). First, we determined the average mRNA levels for each of five classes of genes, representing DNA deletion, no change, and low-, medium-, and high-level amplification (Fig. 4a). For both the

breast cancer cell lines and tumors, average mRNA levels tracked with DNA copy number across all five classes, in a statistically significant fashion ( $P$  values for pair-wise Student's  $t$  tests comparing adjacent classes: cell lines,  $4 \times 10^{-49}$ ,  $1 \times 10^{-49}$ ,  $5 \times 10^{-5}$ ,  $1 \times 10^{-2}$ ; tumors,  $1 \times 10^{-43}$ ,  $1 \times 10^{-214}$ ,  $5 \times 10^{-41}$ ,  $1 \times 10^{-4}$ ). A linear regression of the average  $\log(\text{DNA copy number})$ , for each class, against average  $\log(\text{mRNA level})$  demonstrated that on average, a 2-fold change in DNA copy number was accompanied by 1.4- and 1.5-fold changes in mRNA level for the breast cancer cell lines and tumors, respectively (Fig. 4a, regression line not shown). Second, we characterized the distribution of the 6,095 correlations between DNA copy number and mRNA level, each across the 37 tumor samples (Fig. 4b). The distribution of correlations forms a normal-shaped curve, but with the peak markedly shifted in the positive direction from zero. This shift is statistically significant, as evidenced in a plot of observed vs. expected correlations (Fig. 4c), and reflects a pervasive global influence of DNA copy number alterations on gene expression. Notably, the highest correlations between DNA copy number and mRNA level (the right tail of the distribution in Fig. 4b) comprise both amplified and deleted genes (data not shown). Third, we used a linear regression model to estimate the fraction of all variation measured in mRNA levels among the 37



**Fig. 4.** Genome-wide influence of DNA copy number alterations on mRNA levels. (a) For breast cancer cell lines (gray) and tumor samples (black), both mean-centered mRNA fluorescence ratio (log<sub>2</sub> scale) quantiles (box plots indicate 25th, 50th, and 75th percentile) and averages (diamonds; Y-value error bars indicate standard errors of the mean) are plotted for each of five classes of genes, representing DNA deletion (tumor/normal ratio < 0.8), no change (0.8–1.2), low- (1.2–2), medium- (2–4), and high-level (>4) amplification. *P* values for pair-wise Student's *t* tests, comparing averages between adjacent classes (moving left to right), are  $4 \times 10^{-49}$ ,  $1 \times 10^{-49}$ ,  $5 \times 10^{-53}$ ,  $1 \times 10^{-2}$  (cell lines), and  $1 \times 10^{-48}$ ,  $1 \times 10^{-214}$ ,  $5 \times 10^{-41}$ ,  $1 \times 10^{-4}$  (tumors). (b) Distribution of correlations between DNA copy number and mRNA levels, for 6,095 different human genes across 37 breast tumor samples. (c) Plot of observed versus expected correlation coefficients. The expected values were obtained by randomization of the sample labels in the DNA copy number data set. The line of unity is indicated. (d) Percent variance in gene expression (among tumors) directly explained by variation in gene copy number. Percent variance explained (black line) and fraction of data retained (gray line) are plotted for different fluorescence intensity/background (a rough surrogate for signal/noise) cutoff values. Fraction of data retained is relative to the 1.2 intensity/background cutoff. Details of the linear regression model used to estimate the fraction of variation in gene expression attributable to underlying DNA copy number alteration can be found in the supporting information (see *Estimating the Fraction of Variation in Gene Expression Attributable to Underlying DNA Copy Number Alteration*).

tumors that could be attributed to underlying variation in DNA copy number. From this analysis, we estimate that, overall, about 7% of all of the observed variation in mRNA levels can be explained directly by variation in copy number of the altered genes (Fig. 4d). We can reduce the effects of experimental measurement error on this estimate by using only that fraction of the data most reliably measured (fluorescence intensity/background > 3); using that data, our estimate of the percent variation in mRNA levels directly attributed to variation in gene copy number increases to 12% (Fig. 4d). This still undoubtedly represents a significant underestimate, as the observed variation in global gene expression is affected not only by true variation in the expression programs of the tumor cells themselves, but also by the variable presence of non-tumor cell types within clinical samples.

#### Discussion

This genome-wide, array CGH analysis of DNA copy number alteration in a series of human breast tumors demonstrates the usefulness of defining amplicon boundaries at high resolution (gene-by-gene), and quantitatively measuring amplicon shape, to assist in locating and identifying candidate oncogenes. By analyzing mRNA levels in parallel, we have also discovered that changes in DNA copy number have a large, pervasive, direct effect on global gene expression patterns in both breast cancer

cell lines and tumors. Although the DNA microarrays used in our analysis may display a bias toward characterized and/or highly expressed genes, because we are examining such a large fraction of the genome (approximately 20% of all human genes), and because, as detailed above, we are likely underestimating the contribution of DNA copy number changes to altered gene expression, we believe our findings are likely to be generalizable (but would nevertheless still be remarkable if only applicable to this set of ~6,100 genes).

In budding yeast, aneuploidy has been shown to result in chromosome-wide gene expression biases (13). Two recent studies have begun to examine the global relationship between DNA copy number and gene expression in cancer cells. In agreement with our findings, Phillips *et al.* (14) have shown that with the acquisition of tumorigenicity in an immortalized prostate epithelial cell line, new chromosomal gains and losses resulted in a statistically significant respective increase and decrease in the average expression level of involved genes. In contrast, Platzer *et al.* (15) recently reported that in metastatic colon tumors only ~4% of genes within amplified regions were found more highly (>2-fold) expressed, when compared with normal colonic epithelium. This report differs substantially from our finding that 62% of highly amplified genes in breast cancer exhibit at least 2-fold increased expression. These contrasting findings may reflect methodological differences between the



studies. For example, the study of Platzer *et al.* (15) may have systematically under-measured gene expression changes. In this regard it is remarkable that only 14 transcripts of many thousand residing within unamplified chromosomal regions were found to exhibit at least 4-fold altered expression in metastatic colon cancer. Additionally, their reliance on lower-resolution chromosomal CGH may have resulted in poorly delimiting the boundaries of high-complexity amplicons, effectively overcalling regions with amplification. Alternatively, the contrasting findings for amplified genes may represent real biological differences between breast and metastatic colon tumors; resolution of this issue will require further studies.

Our finding that widespread DNA copy number alteration has a large, pervasive and direct effect on global gene expression patterns in breast cancer has several important implications. First, this finding supports a high degree of copy number-dependent gene expression in tumors. Second, it suggests that most genes are not subject to specific autoregulation or dosage compensation. Third, this finding cautions that elevated expression of an amplified gene cannot alone be considered strong independent evidence of a candidate oncogene's role in tumorigenesis. In our study, fully 62% of highly amplified genes demonstrated moderately or highly elevated expression. This highlights the importance of high-resolution mapping of amplicon boundaries and shape [to identify the "driving" gene(s) within amplicons (16)], on a large number of samples, in addition to functional studies. Fourth, this finding suggests that analyzing

the genomic distribution of expressed genes, even within existing microarray gene expression data sets, may permit the inference of DNA copy number aberration, particularly aneuploidy (where gene expression can be averaged across large chromosomal regions; see Fig. 3 and supporting information). Fifth, this finding implies that a substantial portion of the phenotypic uniqueness (and by extension, the heterogeneity in clinical behavior) among patients' tumors may be traceable to underlying variation in DNA copy number. Sixth, this finding supports a possible role for widespread DNA copy number alteration in tumorigenesis (17, 18), beyond the amplification of specific oncogenes and deletion of specific tumor suppressor genes. Widespread DNA copy number alteration, and the concomitant widespread imbalance in gene expression, might disrupt critical stoichiometric relationships in cell metabolism and physiology (e.g., proteosome, mitotic spindle), possibly promoting further chromosomal instability and directly contributing to tumor development or progression. Finally, our findings suggest the possibility of cancer therapies that exploit specific or global imbalances in gene expression in cancer.

We thank the many members of the P.O.B. and D.B. labs for helpful discussions. J.R.P. was a Howard Hughes Medical Institute Physician Postdoctoral Fellow during a portion of this work. P.O.B. is a Howard Hughes Medical Institute Associate Investigator. This work was supported by grants from the National Institutes of Health, the Howard Hughes Medical Institute, the Norwegian Cancer Society, and the Norwegian Research Council.

1. Kallioniemi, A., Kallioniemi, O. P., Sudar, D., Rutovitz, D., Gray, J. W., Waldman, F. & Pinkel, D. (1992) *Science* 258, 818–821.
2. Kallioniemi, A., Kallioniemi, O. P., Piper, J., Tanner, M., Stokke, T., Chen, L., Smith, H. S., Pinkel, D., Gray, J. W. & Waldman, F. M. (1994) *Proc. Natl. Acad. Sci. USA* 91, 2156–2160.
3. Tirkkonen, M., Tanner, M., Karhu, R., Kallioniemi, A., Isola, J. & Kallioniemi, O. P. (1998) *Genes Chromosomes Cancer* 21, 177–184.
4. Forozan, F., Mahlamaki, E. H., Monni, O., Chen, Y., Veldman, R., Jiang, Y., Gooden, G. C., Ethier, S. P., Kallioniemi, A. & Kallioniemi, O. P. (2000) *Cancer Res.* 60, 4519–4525.
5. Solinas-Toldo, S., Lampel, S., Stilgenbauer, S., Nickolenko, J., Benner, A., Dohner, H., Cremer, T. & Lichter, P. (1997) *Genes Chromosomes Cancer* 20, 399–407.
6. Pinkel, D., Segreaves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W. L., Chen, C., Zhai, Y., *et al.* (1998) *Nat. Genet.* 20, 207–211.
7. Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D. & Brown, P. O. (1999) *Nat. Genet.* 23, 41–46.
8. Perou, C. M., Sortie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., *et al.* (2000) *Nature (London)* 406, 747–752.
9. Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., *et al.* (2001) *Proc. Natl. Acad. Sci. USA* 98, 10869–10874.
10. Schuler, G. D. (1997) *J. Mol. Med.* 75, 694–698.
11. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001) *Nature (London)* 409, 860–921.
12. Fejzo, M. S., Godfrey, T., Chen, C., Waldman, F. & Gray, J. W. (1998) *Genes Chromosomes Cancer* 22, 105–113.
13. Hughes, T. R., Roberts, C. J., Dai, H., Jones, A. R., Meyer, M. R., Slade, D., Burchard, J., Dow, S., Ward, T. R., Kidd, M. J., *et al.* (2000) *Nat. Genet.* 25, 333–337.
14. Phillips, J. L., Hayward, S. W., Wang, Y., Vasselli, J., Pavlovich, C., Padilla-Nash, H., Pezullo, J. R., Ghadimi, B. M., Grossfeld, G. D., Rivera, A., *et al.* (2001) *Cancer Res.* 61, 8143–8149.
15. Platzer, P., Upender, M. B., Wilson, K., Willis, J., Lutterbaugh, J., Noorati, A., Willson, J. K., Mack, D., Ried, T. & Markowitz, S. (2002) *Cancer Res.* 62, 1134–1138.
16. Albertson, D. G., Ylstra, B., Segreaves, R., Collins, C., Dairkee, S. H., Kowbel, D., Kuo, W. L., Gray, J. W. & Pinkel, D. (2000) *Nat. Genet.* 25, 144–146.
17. Li, R., Yerganian, G., Duesberg, P., Kraemer, A., Willer, A., Rausch, C. & Hehlmann, R. (1997) *Proc. Natl. Acad. Sci. USA* 94, 14506–14511.
18. Rasnick, D. & Duesberg, P. H. (1999) *Biochem. J.* 340, 621–630.





# TECHNICAL UPDATE

FROM YOUR LABORATORY SERVICES PROVIDER

## HER-2/neu Breast Cancer Predictive Testing

*Julie Sanford Hanna, Ph.D. and Dan Mornin, M.D.*

EACH YEAR, OVER 182,000 WOMEN in the United States are diagnosed with breast cancer, and approximately 45,000 die of the disease.<sup>1</sup> Incidence appears to be increasing in the United States at a rate of roughly 2% per year. The reasons for the increase are unclear, but non-genetic risk factors appear to play a large role.<sup>2</sup>

Five-year survival rates range from approximately 65%-85%, depending on demographic group, with a significant percentage of women experiencing recurrence of their cancer within 10 years of diagnosis. One of the factors most predictive for recurrence once a diagnosis of breast cancer has been made is the number of axillary lymph nodes to which tumor has metastasized. Most node-positive women are given adjuvant therapy, which increases their survival. However, 20%-30% of patients without axillary node involvement also develop recurrent disease, and the difficulty lies in how to identify this high-risk subset of patients. These patients could benefit from increased surveillance, early intervention, and treatment.

Prognostic markers currently used in breast cancer recurrence prediction include tumor size, histological grade, steroid hormone receptor status, DNA ploidy, proliferative index, and cathepsin D status. Expression of growth factor receptors and over-expression of the HER-2/neu oncogene have also been identified as having value regarding treatment regimen and prognosis.

HER-2/neu (also known as c-erbB2) is an oncogene that encodes a transmembrane glycoprotein that is homologous to, but distinct from, the epidermal growth factor receptor. Numerous studies have indicated that high levels of expression of this protein are associated with rapid tumor growth, certain forms of therapy resistance, and shorter disease-free survival. The gene has been shown to be amplified and/or overexpressed in 10%-30% of invasive breast cancers and in 40%-60% of intraductal breast carcinoma.<sup>3</sup>

There are two distinct FDA-approved methods by which HER-2/neu status can be evaluated: immunohistochemistry (IHC, HercepTest™) and FISH (fluorescent in situ hybridization, PathVysion™ Kit). Both methods can be performed on archived and current specimens. The first method allows visual assessment of the amount of HER-2/neu protein present on the cell membrane. The latter method allows direct quantification of the level of gene amplification present in the tumor, enabling differentiation between low- versus high-amplification. At least one study has demonstrated a difference in

recurrence risk in women younger than 40 years of age for low- versus high-amplified tumors (54.5% compared to 85.7%); this is compared to a recurrence rate of 16.7% for patients with no HER-2/neu gene amplification.<sup>4</sup> HER-2/neu status may be particularly important to establish in women with small ( $\leq 1$  cm) tumor size.

The choice of methodology for determination of HER-2/neu status depends in part on the clinical setting. FDA approval for the Vysis FISH test was granted based on clinical trials involving 1549 node-positive patients. Patients received one of three different treatments consisting of different doses of cyclophosphamide, Adriamycin, and 5-fluorouracil (CAF). The study showed that patients with amplified HER-2/neu benefited from treatment with higher doses of adriamycin-based therapy, while those with normal HER-2/neu levels did not. The study therefore identified a sub-set of women, who because they did not benefit from more aggressive treatment, did not need to be exposed to the associated side effects. In addition, other evidence indicates that HER-2/neu amplification in node-negative patients can be used as an independent prognostic indicator for early recurrence, recurrent disease at any time and disease-related death.<sup>5</sup> Demonstration of HER-2/neu gene amplification by FISH has also been shown to be of value in predicting response to chemotherapy in stage-2 breast cancer patients.

Selection of patients for Herceptin® (Trastuzumab) monoclonal antibody therapy, however, is based upon demonstration of HER-2/neu protein overexpression using HercepTest™. Studies using Herceptin® in patients with metastatic breast cancer show an increase in time to disease progression, increased response rate to chemotherapeutic agents and a small increase in overall survival rate. The FISH assays have not yet been approved for this purpose, and studies looking at response to Herceptin® in patients with or without gene amplification status determined by FISH are in progress.

In general, FISH and IHC results correlate well. However, subsets of tumors are found which show discordant results; i.e., protein overexpression without gene amplification or lack of protein overexpression with gene amplification. The clinical significance of such results is unclear. Based on the above considerations, HER-2/neu testing at SHMC/PAML will utilize immunohistochemistry (HercepTest®) as a screen, followed by FISH in IHC-negative cases. Alternatively, either method may be ordered individually depending on the clinical setting or clinician preference.

## CPT code information

### HER-2/neu via IHC

88342 (including interpretive report)

### HER-2/neu via FISH

88271×2 Molecular cytogenetics, DNA probe, each

88274 Molecular cytogenetics, interphase in situ hybridization, analyze 25-99 cells

88291 Cytogenetics and molecular cytogenetics, interpretation and report

## Procedural Information

Immunohistochemistry is performed using the FDA-approved DAKO antibody kit, Herceptest®. The DAKO kit contains reagents required to complete a two-step immunohistochemical staining procedure for routinely processed, paraffin-embedded specimens. Following incubation with the primary rabbit antibody to human HER-2/neu protein, the kit employs a ready-to-use dextran-based visualization reagent. This reagent consists of both secondary goat anti-rabbit antibody molecules with horseradish peroxidase molecules linked to a common dextran polymer backbone, thus eliminating the need for sequential application of link antibody and peroxidase conjugated antibody. Enzymatic conversion of the subsequently added chromogen results in formation of visible reaction product at the antigen site. The specimen is then counterstained; a pathologist using light-microscopy interprets results.

FISH analysis at SHMC/PAML is performed using the FDA-approved PathVysion™ HER-2/neu DNA probe kit, produced by Vysis, Inc. Formalin fixed, paraffin-embedded breast tissue is processed using routine histological methods, and then slides are treated to allow hybridization of DNA probes to the nuclei present in the tissue section. The Pathvysion™ kit contains two direct-labeled DNA probes, one specific for the alphoid repetitive DNA (CEP 17, spectrum orange) present at the chromosome 17 centromere and the second for the HER-2/neu oncogene located at 17q11.2-12 (spectrum green). Enumeration of the probes allows a ratio of the number of copies of chromosome 17 to the number of copies of HER-2/neu to be obtained; this enables quantification of low versus high amplification levels, and allows an estimate of the percentage of cells with HER-2/neu gene amplification. The clinically relevant distinction is whether the gene amplification is due to increased gene copy number on the two chromosome 17 homologues normally present or an increase in the number of chromosome 17s in the cells. In the majority of cases, ratio equivalents less than 2.0 are indicative of a normal/negative result, ratios of 2.1 and over indicate that amplification is present and to what degree. Interpretation of this data will be performed and reported from the Vysis-certified Cytogenetics laboratory at SHMC.

## References

1. Wingo, P.A., Tong, T., Bolden, S., "Cancer Statistics", 1995;45:1:8-31.
2. "Cancer Rates and Risks", 4th ed., National Institutes of Health, National Cancer Institute, 1996, p. 120.
3. Slamon, D.J., Clark, G.M., Song, S.G., Levin, W.J., Ullrich, A., McGuire, W.L. "Human breast Cancer: Correlation of relapse and survival with amplification of the her-2/neu oncogene". Science, 235:177-182, 1987.
4. Xing, W.R., Gilchrist, K.W., Harris, C.P., Samson, W., Meisner, L.F. "FISH detection of HER-s/neu oncogene amplification in early onset breast cancer". Breast Cancer Res. And Treatment 39(2):203-212, 1996.
5. Press, M.F. Bernstein, L., Thomas, P.A., Meisner, L.F., Zhou, J.Y., Ma, Y., Hung, G., Robinson, R.A., Harris, C., El-Naggar, A., Slamon, D.J., Phillips, R.N., Ross, J.S., Wolman, S.R., Flom, K.J., "Her-2/neu gene amplification characterized by fluorescence in situ hybridization: poor prognosis in node-negative breast carcinomas", J. Clinical Oncology 15(8):2894-2904, 1997.

*Provided for the clients of*

PATHOLOGY ASSOCIATES MEDICAL LABORATORIES  
PACLAB NETWORK LABORATORIES  
TRI-CITIES LABORATORY  
TREASURE VALLEY LABORATORY

*For more information, please contact  
your local representative.*

## Variable expression of the translocated *c-abl* oncogene in Philadelphia-chromosome-positive B-lymphoid cell lines from chronic myelogenous leukemia patients

JAMES B. KONOPKA\*†, STEVEN CLARK\*, JAMI MCLAUGHLIN\*, MASAKUZU NITTA†, YOSHIRO KATO†, ANNABEL STRIFE†, BAYARD CLARKSON†, AND OWEN N. WITTE\*§

\*Department of Microbiology and Molecular Biology Institute, University of California, Los Angeles, 405 Hilgard Avenue, Los Angeles, CA 90024; and †The Laboratory of Hematopoietic Cell Kinetics and The Laboratory of Cancer Genetics and Cytogenetics, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021

Communicated by Michael Potter, February 10, 1986

**ABSTRACT** The consistent cytogenetic translocation of chronic myelogenous leukemia (the Philadelphia chromosome, Ph<sup>1</sup>) has been observed in cells of multiple hematopoietic lineages. This translocation creates a chimeric gene composed of breakpoint-cluster-region (*bcr*) sequences from chromosome 22 fused to a portion of the *abl* oncogene on chromosome 9. The resulting gene product (P210<sup>c-abl</sup>) resembles the transforming protein of the Abelson murine leukemia virus in its structure and tyrosine kinase activity. P210<sup>c-abl</sup> is expressed in Ph<sup>1</sup>-positive cell lines of myeloid lineage and in clinical specimens with myeloid predominance. We show here that Epstein-Barr virus-transformed B-lymphocyte lines that retain Ph<sup>1</sup> can express P210<sup>c-abl</sup>. The level of expression in these B-cell lines is generally lower and more variable than that observed for myeloid lines. Protein expression is not related to amplification of the *abl* gene but to variation in the level of *bcr-abl* mRNA produced from a single Ph<sup>1</sup> template.

Chronic myelogenous leukemia (CML) is a disease of the pluripotent stem cell (1). In greater than 95% of patients, the leukemic cells contain the cytogenetic marker known as the Philadelphia chromosome, or Ph<sup>1</sup> (2). This reciprocal translocation event between the long arms of chromosomes 9 and 22 has been used as a disease-specific marker for diagnosis and evaluation of therapy. Multiple hematopoietic lineages, including myeloid and B-lymphoid, contain Ph<sup>1</sup> in early or chronic phase, as well as in the more acute accelerated and blast crisis phases of the disease.

One molecular consequence of Ph<sup>1</sup> is the translocation of the chromosomal arm containing the *c-abl* gene on chromosome 9 into the middle of the breakpoint-cluster region (*bcr*) gene on chromosome 22 (3-6). Although the precise translocation breakpoints are variable, an RNA-splicing mechanism generates a very similar 8-kilobase (kb) mRNA in each case (5-9). The hybrid *bcr-abl* message encodes a structurally altered form of the *abl* oncogene product, called P210<sup>c-abl</sup> (10-13), with an amino-terminal segment derived from a portion of the exons of *bcr* on chromosome 22 and a carboxyl-terminal segment derived from a major portion of the exons of the *c-abl* gene on chromosome 9. The chimeric structure of *bcr-abl* and the resulting P210<sup>c-abl</sup> is similar to the structure of the Abelson murine leukemia virus *gag-abl* genome and resulting P160<sup>v-abl</sup> transforming gene product. Both proteins have very similar tyrosine kinase activities (10, 11, 14) which can be distinguished by their relative stability to denaturing detergents and by their ATP requirements from the recently described tyrosine kinase activity of the *c-abl* gene product (15).

In concert with structural modification of the amino-terminal portion of the *abl* gene, increased level of expression has been implicated in activation of *c-abl* oncogenic potential. Myeloid and erythroid cell lines and clinical samples derived from acute-phase CML patients contain about 10-fold higher levels of the 8-kb *bcr-abl* mRNA and P210<sup>c-abl</sup> than the *c-abl* mRNA forms (6 and 7 kb) and P145<sup>c-abl</sup> gene product (5, 8, 9, 11). The higher level of expression of the chimeric *bcr-abl* message in acute-phase cells is not likely to be solely due to the presence of the *bcr* promoter sequences at the 5' end of the gene, since the normal 4.5-kb and 6.7-kb *bcr*-encoded mRNA species are expressed at an even lower level than the normal *c-abl* messages (5, 6).

We have analyzed a series of Epstein-Barr virus-immortalized B-lymphoid cell lines derived from CML patients (16). With such *in vitro* clonal cell lines, we can evaluate whether the presence of Ph<sup>1</sup> always results in synthesis of the chimeric *bcr-abl* message and protein, and whether the quantitative expression varies for cells of B-lymphoid lineage as compared to previously examined myeloid cell lines. Our results show that cell lines that retain Ph<sup>1</sup> do express *bcr-abl* message and protein, but that the level is generally lower and more variable than previously seen for myeloid cell lines. The demonstration that the Ph<sup>1</sup> chromosomal template can vary in its level of expression of P210<sup>c-abl</sup> suggests that secondary mechanisms, beyond the translocation itself, contribute to the regulation of the *bcr-abl* gene in different cell types or subclones that derive from the affected stem cell.

### MATERIALS AND METHODS

**Cells and Cell Labelings.** Epstein-Barr virus-transformed B-lymphoid cell lines were established from peripheral blood samples of chronic- and acute-phase CML patients as reported (16). The cell lines are designated according to patient number, karyotype, and lineage. For example, SK-CML7Bt(9,22)-33 refers to CML patient 7, B-lymphoid cell line, 9;22 translocation (Ph<sup>1</sup>), cell line 33; and SK-CML7BN-2 refers to B-cell line 2 with a normal karyotype derived from the same patient. Repeat karyotype analysis was performed to verify the retention of Ph<sup>1</sup> just prior to analysis for *abl* protein and RNA. Cells were maintained in RPMI 1640 medium with 20% fetal bovine serum. We have not observed any consistent pattern of *in vitro* growth rate that correlates to the stage of disease at the time of transformation with Epstein-Barr virus. Cells (1.5 × 10<sup>7</sup>) were washed twice with Dulbecco's modified Eagle's medium lacking phosphate and

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: *bcr*, breakpoint-cluster region; CML, chronic myelogenous leukemia; kb, kilobase(s).

†Present address: Department of Genetics, University of Washington, Seattle, WA 98195.

§To whom correspondence should be addressed.

supplemented with 5% dialyzed fetal bovine serum. Cells were then resuspended in 2 ml of the minimal medium. Labeling was started with the addition of [ $^{32}$ P]orthophosphate (1 mCi/ml; ICN; 1 Ci = 37 GBq) and continued at 37°C for 3–4 hr.

**Immunoprecipitation and Immunoblotting.** Immunoprecipitations were carried out as described (10). Cells ( $1.5 \times 10^7$ ) were washed with phosphate-buffered saline and extracted with 3–5 ml of phosphate lysis buffer (1% Triton X-100/0.1 NaDodSO<sub>4</sub>/0.5% deoxycholate/10 mM Na<sub>2</sub>HPO<sub>4</sub>, pH 7.5/100 mM NaCl) with 5 mM EDTA and 5 mM phenylmethylsulfonyl fluoride. Extracts were clarified by centrifugation and precipitated with normal or rabbit anti-*abl* sera (anti-pEX-2 or anti-pEX-5) (17). The precipitated proteins were electrophoresed in a NaDodSO<sub>4</sub>/8% polyacrylamide gel.  $^{32}$ P-labeled proteins were detected by autoradiography. Alternatively, *abl* proteins were detected by immunoblotting. Extracts from unlabeled cells were clarified, and proteins were concentrated by immunoprecipitation with rabbit antisera against *abl*-encoded proteins [anti-pEX-2 and anti-pEX-5 combined (17)] and then fractionated in 8% acrylamide gels. The proteins were transferred from the gel to nitrocellulose filters, using protease-facilitated transfer (18). The *abl*-encoded proteins were detected using murine monoclonal antibodies as a probe and peroxidase-conjugated goat anti-mouse second stage antibody (Bio-Rad) for development. Rabbit antisera and mouse monoclonal antibodies to *abl* proteins were prepared using bacterially expressed regions of the *v-abl* protein as immunogens (17, 19). Anti-pEX-2 antibodies react with the internal tyrosine kinase domain and anti-pEX-5 antibodies react with the carboxyl-terminal segment of the *abl* proteins.

**RNA Analysis.** RNA was extracted from  $10^8$  cells by the NaDodSO<sub>4</sub>/urea/phenol method (20). Polyadenylated RNA was purified by oligo(dT) affinity chromatography. Samples were electrophoresed in a 1% agarose/formaldehyde gel and transferred to nitrocellulose. *abl* RNA species were detected by hybridization with a nick-translated *v-abl* fragment probe (21).

**DNA Analysis.** DNA was prepared from  $5 \times 10^7$  cells of each cell line and processed for Southern blots with a *v-abl* probe as described (21).

## RESULTS

**Variable Levels of P210<sup>c-abl</sup> Are Detected in Ph<sup>1</sup>-Positive Cell Lines.** Ph<sup>1</sup>-positive and Ph<sup>1</sup>-negative, Epstein-Barr virus-transformed B-lymphocyte cell lines derived from the same patient were examined for P210<sup>c-abl</sup> synthesis by immunoprecipitation of [ $^{32}$ P]orthophosphate-labeled cell extracts with anti-*abl* sera (Fig. 1). The normal *c-abl* protein P145<sup>c-abl</sup> was detected at a similar level in multiple Ph<sup>1</sup>-positive and Ph<sup>1</sup>-negative cell lines. P210<sup>c-abl</sup> was only detected in the Ph<sup>1</sup>-positive cell lines because the *bcr-abl* chimeric gene which encodes P210<sup>c-abl</sup> resides on the Ph<sup>1</sup> (4, 5, 11, 13). The level of P210<sup>c-abl</sup> was about 4- to 5-fold higher than the level of P145<sup>c-abl</sup> in the SK-CML7Bt-33 cell line (Fig. 1A, +). The Ph<sup>1</sup>-positive erythroid-progenitor cell line K562 (C) showed a level of P210<sup>c-abl</sup> about 10-fold higher than P145<sup>c-abl</sup>. However, the level of P210<sup>c-abl</sup> was about one-fifth that of P145<sup>c-abl</sup> in the Ph<sup>1</sup>-positive SK-CML16Bt-1 cell line (Fig. 1B, +). Comparison of different autoradiographic exposures roughly indicated that the level of P210<sup>c-abl</sup> varies over a 20-fold range between these Ph<sup>1</sup>-positive B-cell lines. Analysis of four additional Ph<sup>1</sup>-positive B-cell lines demonstrated that the level of P210<sup>c-abl</sup> fell into two general classes; some cell lines had a level of P210<sup>c-abl</sup> similar to SK-CML7Bt-33 and others had the low level similar to SK-CML16Bt-1 (Table 1). This differs from previous studies with Ph<sup>1</sup>-positive myeloid cell lines and patient samples derived from acute-

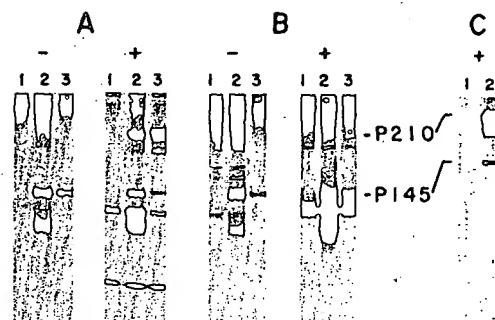


Fig. 1. Detection of variable levels of P210<sup>c-abl</sup> in Ph<sup>1</sup>-positive B-cell lines. Production of P145<sup>c-abl</sup> and P210<sup>c-abl</sup> in Epstein-Barr virus-transformed B-cell lines derived from a blast-crisis (A) and a chronic-phase (B) CML patient was examined by metabolic labeling with [ $^{32}$ P]orthophosphate and immunoprecipitation. Ph<sup>1</sup>-negative (–) and Ph<sup>1</sup>-positive (+) cell lines derived from each patient were analyzed. The Ph<sup>1</sup>-negative cell line in A, – is SK-CML7Bt-33 and in B, – is SK-CML16Bt-1. The Ph<sup>1</sup>-positive cell line in A, + is SK-CML7Bt-33 and in B, + is SK-CML16Bt-1. The K562 cell line, a Ph<sup>1</sup>-positive erythroid progenitor cell line spontaneously derived from a blast-crisis patient (33), is represented in C. Cells ( $1.5 \times 10^7$ ) were metabolically labeled with 2 mCi of [ $^{32}$ P]orthophosphate for 3–4 hr and then were extracted and clarified by centrifugation. Samples were immunoprecipitated with control normal serum (lanes 1), anti-pEX-2 (lanes 2), or anti-pEX-5 (lanes 3) and analyzed by NaDodSO<sub>4</sub>/8% PAGE followed by autoradiography with an intensifying screen (3 days for A and C, 10 days for B).

phase CML patients, in which P210<sup>c-abl</sup> was detected at a 10-fold higher level than P145<sup>c-abl</sup> (refs. 10 and 11; Table 1). There was no large difference in level of chimeric mRNA and P210<sup>c-abl</sup> expressed in four myeloid/erythroid-lineage Ph<sup>1</sup>-positive cell lines (K562, EM2, EM3, CML22, and BV173; refs. 9 and 11), despite a 4- to 5-fold amplification of *abl*-related sequences in the K562 cell line.

Detection of different levels of P210<sup>c-abl</sup> in Fig. 1 could be due to decreased phosphorylation of P210<sup>c-abl</sup>, a lower level of P210<sup>c-abl</sup> synthesis, or altered stability of the protein. To help distinguish among these possibilities, the steady-state level of P210<sup>c-abl</sup> in the cell lines was assayed by immunoblotting. The results show that SK-CML7Bt-33 (Fig. 2A, +) had a higher level of P210<sup>c-abl</sup> than P145, similar to the results with metabolic labeling (Fig. 1). We did not detect P210<sup>c-abl</sup> by immunoblotting with  $2 \times 10^7$  cells of line SK-CML8Bt-3 (Fig. 2B, +). Reconstruction experiments using dilutions of cell extracts showed that we could detect about 5–10% the level of P210<sup>c-abl</sup> expressed in the K562 cell line (data not shown). We infer that the steady-state level of P210<sup>c-abl</sup> in SK-CML8Bt-3 is lower than the level in SK-CML7Bt-33 by a factor of at least 10. The level of P210<sup>c-abl</sup> detected in these assays correlated with the amount of P210<sup>c-abl</sup> tyrosine kinase activity that could be detected *in vitro* (data not shown).

**Different Levels of P210<sup>c-abl</sup> Are Reflected in the Amount of Stable *bcr-abl* mRNA.** To identify the basis for detection of variable levels of P210<sup>c-abl</sup>, we examined the production of the *abl* RNA. RNA blot hybridization analysis using a *v-abl* probe (Fig. 3) showed that the normal 6- and 7-kb *c-abl* mRNAs were present at a similar level in Ph<sup>1</sup>-positive and -negative cell lines derived from different patients. However, the 8-kb mRNA that encodes P210<sup>c-abl</sup> was detected at a 10-fold higher level in SK-CML7Bt-33 (Fig. 3A, +) than in SK-CML16Bt-1 (B, +), which correlated with the relative level of P210<sup>c-abl</sup> detected in each cell line. Analysis of additional cell lines demonstrated that the level of 8-kb RNA directly correlated with the level of P210<sup>c-abl</sup> (Table 1). The variation in level of 8-kb RNA detected in these cell lines was not due to loss or gain of Ph<sup>1</sup>, because cytogenetic analysis confirmed the presence of Ph<sup>1</sup> in these cell lines (ref. 16 and

Table 1. Relative levels of *bcr-abl* expression in Epstein-Barr virus-immortalized B-cell lines and myeloid CML lines

Cell line*	CML phase†	Ph <sup>+</sup> ‡	P210§	8-kb mRNA¶
SK-CML7BN-2	BC	-	-	-
SK-CML8BN-10	Chronic	-	-	-
SK-CML8BN-12	Chronic	-	-	-
SK-CML16BN-1	Chronic	-	-	-
SK-CML35BN-1	Chronic	-	-	-
SK-CML7B5-33	BC	+	+++	+++
SK-CML21Bt-1	Acc	+	+++	+++
SK-CML21Bt-6	Acc	+	+++	+++
SK-CML8Bt-3	Chronic	+	+	±
SK-CML16Bt-1	Chronic	+	+	+
SK-CML35Bt-2	Chronic	+	+	+
K562	BC	+	+++++	+++++
BV173	BC	+	+++++	+++++
EM2	BC	+	+++++	+++++

\*Cell lines derived from CML patients by transformation with Epstein-Barr virus as described (16). Names of cell lines indicate patient number and Ph<sup>+</sup> status: SK-CML7Bt indicates a cell line derived from patient 7 that carries the 9;22 Ph<sup>+</sup> translocation; N indicates a normal karyotype. Myeloid-erythroid cell lines (K562, EM2, and BV173) are described in previous publications (9, 11, 22, 33).

†Status of patient at the time cell line was derived. BC, blast crisis; Acc, accelerated phase.

‡Presence (+) or absence (-) of Ph<sup>+</sup> as demonstrated by karyotypic or Southern blot analysis.

§P210<sup>c-abl</sup> detected as described in legend to Fig. 1. B-cell lines derived from blast-crisis and accelerated-phase patients had levels of P210 3- to 5-fold higher (++++) than levels of P145. Chronic-phase-derived cell lines had P210 levels lower than or just equivalent (+) to the level of P145. Myeloid and erythroid lines had levels of P210 5- to 10-fold higher than P145 (+++++).

¶Eight-kilobase *bcr-abl* mRNA detected as described in legend to Fig. 2. Symbols: ±, borderline detectable; +++++, level of 8-kb mRNA 5- to 10-fold higher than that of the 6- and 7-kb *c-abl* mRNA species; +++, level of 8-kb mRNA 3- to 5-fold higher than that of the 6- and 7-kb species; +, a level approximately equivalent to that of the 6- and 7-kb messages.

data not shown). There was no difference in the copy number of *abl*-related sequences as judged by Southern blot analysis (Fig. 4). Only the K562 cell line control showed an amplification of *abl* sequences, as previously reported (22, 23). These combined data suggest that differential *bcr-abl* mRNA expression from a single gene template is responsible for the variable levels of P210<sup>c-abl</sup> detected. This could be mediated

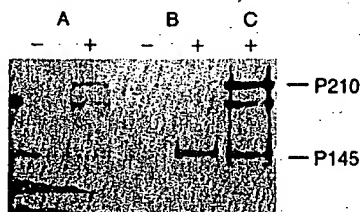


FIG. 2. Analysis of steady-state *abl* protein levels by immunoblotting. Cell extracts prepared from  $2 \times 10^7$  cells of lines SK-CML7BN-2 (A, -), SK-CML7Bt-33 (A, +), SK-CML8BN-10 (B, -), and SK-CML8Bt-3 (B, +) were concentrated by immunoprecipitation with anti-pEX-2 plus anti-pEX-5. Samples were then electrophoresed in a NaDodSO<sub>4</sub>/8% polyacrylamide gel and transferred to nitrocellulose, using protease-facilitated transfer (18). *abl* proteins were detected using a mixture of two monoclonal antibodies directed against the pEX-2 and pEX-5 *abl*-protein fragments produced in bacteria (19) as a probe and a peroxidase-conjugated goat anti-mouse second-stage antibody (Bio-Rad) for development.

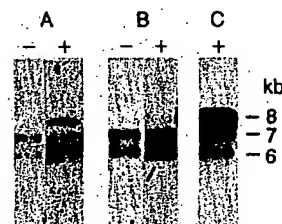


FIG. 3. Comparison of *abl* RNA levels in Ph<sup>+</sup>-positive and -negative B-cell lines. The levels of the normal 6- and 7-kb *c-abl* RNAs and the 8-kb *bcr-abl* RNA were analyzed by blot hybridization using a *v-abl* probe. RNA was extracted from Ph<sup>+</sup>-negative lines SK-CML7BN-2 (A, -) and SK-CML16BN-1 (B, -), from Ph<sup>+</sup>-positive lines SK-CML6Bt-33 (A, +) and SK-CML16Bt-3 (B, +), and from line K562 (C, +) by the NaDodSO<sub>4</sub>/urea/phenol method (20). Polyadenylated RNA was purified by oligo(dT) affinity chromatography, and 15  $\mu$ g of each sample was electrophoresed in a 1% agarose/formaldehyde gel and then transferred to nitrocellulose. The blotted RNAs were hybridized with a nick-translated *v-abl* fragment probe (21) and then autoradiographed for 4 days.

by factors influencing the transcription rate of the *bcr-abl* gene or the stability of the mRNA.

## DISCUSSION

Several lines of evidence suggest that formation of Ph<sup>+</sup> is not the primary event that affects the stem cell in CML. Patients have been identified that present with the clinical picture of CML but only later develop Ph<sup>+</sup> (1). This observation, coupled with studies of *G6PD* (glucose-6-phosphate dehydrogenase)-heterozygous females with CML that demonstrate stem-cell clonality by isozyme analysis among cell

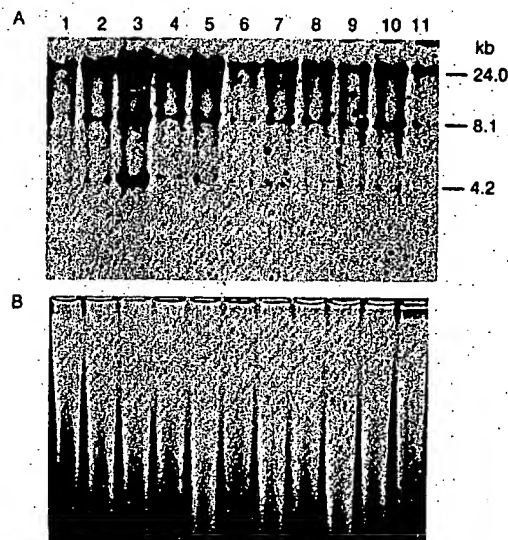


FIG. 4. Southern blot analysis of *abl* sequences in Ph<sup>+</sup>-positive and -negative B-cell lines. High molecular weight DNA (15  $\mu$ g) was digested with restriction endonuclease *Bam*HI, separated in a 0.8% agarose gel, and then transferred to nitrocellulose. The blotted DNA fragments were hybridized with a nick-translated, 2.4-kb *Bgl* II *v-abl* fragment ( $1.5 \times 10^8$  cpm/ $\mu$ g; ref. 21) and exposed for 4 days. (A) Autoradiograph of *abl*-specific fragments in cell lines HL-60 (lane 1), EM2 (lane 2), K562 (lane 3), SK-CML7Bt-33 (lane 4), SK-CML8Bt-3 (lane 5), SK-CML16Bt-1 (lane 6), SK-CML21Bt-6 (lane 7), SK-CML35Bt-2 (lane 8), SK-CML7BN-2 (lane 9), SK-CML8BN-2 (lane 10), and SK-CML35BN-1 (lane 11). (B) Ethidium bromide staining of agarose gel prior to transfer to nitrocellulose, showing the level of variation in amount of DNA loaded per lane.

populations that lack the Ph<sup>1</sup> marker, supports a secondary or complementary role for Ph<sup>1</sup> in the progression of the disease (24, 25). This chromosome marker is found in chronic, accelerated, and blast-crisis phases of the disease. It is likely that Ph<sup>1</sup> confers some growth advantage, since cells with the marker chromosome eventually predominate the marrow and peripheral blood even in chronic phase. During the phase of blast crisis, many patients develop additional chromosome abnormalities, including duplication of Ph<sup>1</sup>, a variety of trisomies, and complex translocations (26). This is suggestive evidence for Ph<sup>1</sup> being a necessary but not sufficient genetic change for the full evolution of the disease.

The realization that one molecular result of Ph<sup>1</sup> is the generation of a chimeric *bcr-abl* protein with functional characteristics and structure analogous to the *gag-abl* transforming protein of the Abelson murine leukemia virus strengthens the argument for an important role of Ph<sup>1</sup> in the pathogenesis of CML. Although the Abelson virus is generally considered a rapidly transforming retrovirus, its effects can range from overcoming growth factor requirements, to cellular lethality, to induction of highly oncogenic tumors in a number of hematopoietic cell lineages (27, 28). Even in the transformation of murine cell targets, there are several lines of evidence that suggest that the growth-promoting activity of the *v-abl* gene product is complemented by further cellular changes in the production of the malignant-cell phenotype (29–31).

The regulation of *bcr-abl* gene expression is complex because the 5' end of the gene is derived from the non-*abl* sequences, *bcr*, normally found on chromosome 22 (6). The level of stable message for the normal *bcr* gene and the normal *abl* gene are both much lower than the level of the *bcr-abl* message and protein from cell lines and clinical specimens derived from myeloid blast-crisis patients (5, 6, 11). Therefore, the high level of *bcr-abl* expression cannot simply be attributed to the regulatory sequences associated with *bcr*. Possibly, creation of the chimeric gene disrupts the normal regulatory sequences and results in a higher level of expression. Variation in *bcr-abl* expression may result from secondary changes in the structure of the chimeric gene or function of *trans*-acting factors that occur during evolution of the disease. Our analysis of P210<sup>c-abl</sup> and the 8-kb mRNA in Epstein-Barr virus-transformed Ph<sup>1</sup>-positive B-cell lines demonstrates that stable message and protein levels from the *bcr-abl* gene can vary over a wide range. This variation does not result from a change in the number of *bcr-abl* templates secondary to gene amplification but more likely from changes in either transcription rate or mRNA stability. We suspect this range of *bcr-abl* expression is not limited to lymphoid cells. Analysis of peripheral blood leukocytes derived from an unusual CML patient who has been in chronic phase with myeloid predominance for 16 years showed a level of P210<sup>c-abl</sup> one-fifth that of P145<sup>c-abl</sup>, as detected by metabolic labeling with [<sup>32</sup>P]orthophosphate and immunoprecipitation (S.C., O.N.W., and P. Greenberg, unpublished observations). Lower levels of expression of the chimeric mRNA have been demonstrated in clinical samples from chronic-phase CML patients compared to acute-phase CML patients (9). Others have reported chronic-phase patients with variable but, in some cases, relatively high levels of the *bcr-abl* mRNA (32). The sampling variation and the heterogeneous mixture of cell types in clinical samples complicate such analyses. Further work is needed to evaluate whether there is a defined change in P210<sup>c-abl</sup> expression during the progression of CML. It is interesting to note that among the limited sample of Ph<sup>1</sup>-positive B-cell lines we have examined (Table 1), we have seen higher levels of P210<sup>c-abl</sup> in those derived from patients at more advanced stages of the disease.

It will be important to search for cell-type-specific mechanisms that might regulate expression of *bcr-abl* from Ph<sup>1</sup>.

We thank Bonnie Hechinger and Carol Crookshank for excellent secretarial assistance and Margaret Newman for excellent technical assistance. This work was supported by grants from the National Institutes of Health (to O.N.W. and B.C.). J.B.K. was supported as a predoctoral fellow on the Public Health Service Cellular and Molecular Biology Training Grant GM07185. S.C. is a postdoctoral fellow of the Leukemia Society of America.

- Champlin, R. E. & Golde, D. W. (1985) *Blood* 65, 1039–1047.
- Rowley, J. D. (1973) *Nature (London)* 243, 290–291.
- Heisterkamp, N., Stephenson, J. R., Groffen, J., Hansen, P. F., de Klein, A., Bartram, C. R. & Grosveld, G. (1983) *Nature (London)* 306, 239–242.
- Bartram, C. R., de Klein, A., Hagemeijer, A., van Agthoven, T., van Kessel, A. G., Bootsma, D., Grosveld, G., Ferguson-Smith, M. A., Davies, T., Stone, M., Heisterkamp, N., Stephenson, J. R. & Groffen, J. (1983) *Nature (London)* 306, 277–280.
- Shtivelman, E., Lifshitz, B., Gale, R. P. & Canaani, D. (1985) *Nature (London)* 315, 550–554.
- Heisterkamp, N., Stam, K. & Groffen, J. (1985) *Nature (London)* 315, 758–761.
- Groffen, J., Stephenson, J. R., Heisterkamp, N., de Klein, A., Bartram, C. R. & Grosveld, G. (1984) *Cell* 36, 93–99.
- Gale, R. P. & Canaani, E. (1984) *Proc. Natl. Acad. Sci. USA* 81, 5648–5652.
- Collins, S., Kubonishi, L., Miyoshi, I. & Groudine, M. T. (1984) *Science* 225, 72–74.
- Konopka, J. B., Watanabe, S. M. & Witte, O. N. (1984) *Cell* 7, 1035–1042.
- Konopka, J. B., Watanabe, S. M., Singer, J., Collins, S. & Witte, O. N. (1985) *Proc. Natl. Acad. Sci. USA* 82, 1810–1814.
- Kloetzer, W., Kurzrock, R., Smith, L., Talpaz, M., Spiller, M., Gutterman, J. & Arlinghaus, R. (1985) *Virology* 140, 230–238.
- Kozbor, D., Giallongo, A., Sierzega, M. E., Konopka, J. B., Witte, O. N., Showe, L. C. & Croce, C. M. (1985) *Nature (London)*, in press.
- Davis, R. L., Konopka, J. B. & Witte, O. N. (1985) *Mol. Cell Biol.* 5, 204–213.
- Konopka, J. B. & Witte, O. N. (1985) *Mol. Cell Biol.* 5, 3116–3123.
- Nitta, M., Kato, Y., Strife, A., Wachter, M., Fried, J., Perez, A., Jhanwar, S., Duigou, R., Chaganti, R. S. K. & Clarkson, B. (1985) *Blood* 66, 1053–1061.
- Konopka, J. B., Davis, J. L., Watanabe, S. M., Ponticelli, A. S., Schiff-Maker, L., Rosenberg, N. & Witte, O. N. (1984) *Virology* 51, 223–232.
- Gibson, W. (1981) *Anal. Biochem.* 118, 1–3.
- Schiff-Maker, L., Konopka, J. B., Clark, S., Witte, O. N. & Rosenberg, N. (1986) *J. Virol.* 57, 1182–1186.
- Schwartz, R. C., Sonenshein, G. E., Bothwell, A. & Gelfand, M. L. (1981) *J. Immunol.* 126, 2104–2108.
- Goff, S. P., Gilboa, E., Witte, O. N. & Baltimore, D. (1980) *Cell* 22, 777–785.
- Collins, S. J. & Groudine, M. T. (1983) *Proc. Natl. Acad. Sci. USA* 80, 4813–4817.
- Selden, J. R., Emanuel, B. S., Wang, E., Cannizzaro, L., Palumbo, A., Erikson, J., Nowell, P. C., Rovera, G. & Croce, C. M. (1983) *Proc. Natl. Acad. Sci. USA* 80, 7289–7292.
- Fialkow, P. J., Martin, P. J., Najfeld, V., Penfold, G. K., Jacobson, R. J. & Hansen, J. A. (1981) *Blood* 58, 158–163.
- Martin, P. J., Najfeld, V. & Fialkow, P. J. (1982) *Can. Gen. Cytogenet.* 6, 359–368.
- Rowley, J. D. (1980) *Annu. Rev. Genet.* 14, 17–40.
- Whitlock, C. A. & Witte, O. N. (1984) *Adv. Immunol.* 37, 74–98.
- Pierce, J. H., Di Fiore, P. P., Aaronson, S. A., Potter, M., Pumphrey, J., Scott, A. & Ihle, J. N. (1985) *Cell* 41, 685–693.
- Whitlock, C. A., Ziegler, S. & Witte, O. N. (1983) *Mol. Cell Biol.* 3, 596–604.
- Wolf, D., Harris, N. & Rotter, V. (1984) *Cell* 38, 119–126.
- Klein, G. & Klein, E. (1985) *Nature (London)* 315, 190–195.
- Stam, K., Jr., Heisterkamp, N., Grosveld, G., de Klein, A., Verma, R., Coleman, M., Dosik, H. & Groffen, J. (1985) *N. Engl. J. Med.* 313, 1429–1433.
- Lozzio, C. B. & Lozzio, B. B. (1975) *Blood* 45, 321–334.



## Analysis of Genomic and Proteomic Data Using Advanced Literature Mining

Yanhui Hu, Lisa M. Hines, Haifeng Weng, Dongmei Zuo, Miguel Rivera,  
Andrea Richardson, and Joshua LaBaer\*

*Institute of Proteomics, Harvard Medical School—BCMP, 240 Longwood Avenue, Boston, Massachusetts 02115*

Received March 13, 2003

High-throughput technologies, such as proteomic screening and DNA micro-arrays, produce vast amounts of data requiring comprehensive analytical methods to decipher the biologically relevant results. One approach would be to manually search the biomedical literature; however, this would be an arduous task. We developed an automated literature-mining tool, termed MedGene, which comprehensively summarizes and estimates the relative strengths of all human gene–disease relationships in Medline. Using MedGene, we analyzed a novel micro-array expression dataset comparing breast cancer and normal breast tissue in the context of existing knowledge. We found no correlation between the strength of the literature association and the magnitude of the difference in expression level when considering changes as high as 5-fold; however, a significant correlation was observed ( $r = 0.41$ ;  $p = 0.05$ ) among genes showing an expression difference of 10-fold or more. Interestingly, this only held true for estrogen receptor (ER) positive tumors, not ER negative. MedGene identified a set of relatively understudied, yet highly expressed genes in ER negative tumors worthy of further examination.

**Keywords:** bioinformatics • micro-array • text mining • gene–disease association • breast cancer

### Introduction

At its current pace, the accumulation of biomedical literature outpaces the ability of most researchers and clinicians to stay abreast of their own immediate fields, let alone cover a broader range of topics. For example, to follow a single disease, e.g., breast cancer, a researcher would have had to scan 130 different journals and read 27 papers per day in 1999.<sup>1</sup> This problem is accentuated with high-throughput technologies such as DNA micro-arrays and proteomics, which require the analysis of large datasets involving thousands of genes, many of which are unfamiliar to a particular researcher. In any microarray experiment, thousands of genes may demonstrate statistically significant expression changes, but only a fraction of these may be relevant to the study. The ability to interpret these datasets would be enhanced if they could be compared to a comprehensive summary of what is known about all genes. Thus, there is a need to summarize existing knowledge in a format that allows for the rapid analysis of associations between genes and diseases or other specific biological concepts.

One solution to this problem is to compile structured digital resources, such as the Breast Cancer Gene Database<sup>1</sup> and the Tumor Gene Database.<sup>2</sup> However, as these resources are hand-curated, the labor-intensive review process becomes a rate-limiting step in the growth of the database. As a result, these

databases have a limited scale and the genes are not selected in a systematic fashion.

An alternative approach is automated text mining; a method which involves automated information extraction by searching documents for text strings and analyzing their frequency and context. This approach has been used successfully in several instances for biological applications. In most cases, it has been applied to extract information about the relationships or interactions that proteins or genes have with one another, in the literature or by functional annotation.<sup>3–7</sup> Thus far, few publications have applied text-mining to examine the global relationships between genes and diseases. Perez-Iratxeta et al. automatically examined the GO (Gene Ontology) annotation of genes and their predicted chromosomal locations in order to identify genes linked to inherited disorders.<sup>8</sup>

To obtain a more global understanding of disease development, it would be valuable to incorporate information regarding all possible gene–disease relationships, including biochemical, physiological, pharmacological, epidemiological, as well as genetic. This information would enable comprehensive comparisons between large experimental datasets and existing knowledge in the literature. This would accomplish two things. First, it would serve to validate experiments by demonstrating that known responses occur as predicted. Second, it would rapidly highlight which genes are corroborated by the literature and which genes are novel in a given context. We have utilized a computational approach to literature mining to produce a

\* To whom correspondence should be addressed: jlabacr@hms.harvard.edu.

## research articles

comprehensive set of gene-disease relationships. In addition, we have developed a novel approach to assess the strength of each association based on the frequency of citation and co-citation. We applied this tool to help interpret the data from a large micro-array gene expression experiment comparing normal and cancerous breast tissue.

### Methods

**MedGene Database.** MedGene is a relational database, storing disease and gene information from NCBI, text mining results, statistical scores, and hyperlinks to the primary literature. MedGene has a web-based user interface for users to query the database (<http://hipseq.med.harvard.edu/MedGene/>).

**Text Mining Algorithms.** MeSH files were downloaded from the MeSH web site at NLM (National Library of Medicine) (<http://www.nlm.nih.gov/mesh/meshhome.html>) and human disease categories were selected. LocusLink files were downloaded from the LocusLink web site at NCBI (<http://www.ncbi.nlm.nih.gov/LocusLink/>). Official/preferred gene symbol, official/preferred gene name, and gene alternative symbols and names, all relevant annotations and URLs for each LocusLink record, were collected. Gene search terms were used for literature searching and included all qualified gene names, gene symbols, and gene family terms. Primary gene keys, predominantly qualified gene family terms and gene official/preferred symbols, were used to index Medline records. If the official/preferred gene symbols did not meet the standards to be an index, then qualified gene official/preferred names were used. A local copy of Medline records (up to July, 2002) was pre-selected.

A JAVA module examined the MeSH terms and then indexed each Medline record with the appropriate disease terms. A separate JAVA module was used to examine the titles and abstracts for gene search terms and then to index the gene-related Medline records with the relevant primary gene key(s).

**Statistical Methods.** For every gene and disease pair, we counted records that were indexed for both gene and disease (double positive hits), for disease only (disease single hits), for gene only (gene single hits), and for neither gene nor disease (double negative hits) to generate a  $2 \times 2$  contingency table. On the basis of the contingency table-framework, we applied different statistical methods to estimate the strength of gene-disease relationships and evaluated the results. These methods included chi-square analysis, Fisher's exact probabilities, relative risk of gene, and relative risk of disease<sup>10</sup> (<http://hipseq.med.harvard.edu/MedGene/>). In addition, we computed the "product of frequency", which is the product of the proportion of disease/gene double hits to disease single hits and the proportion of disease/gene double hits to gene single hits. To obtain a normal distribution, we transformed all the statistical scores using the natural logarithm. We selected the log of the product of frequency (LPF) to validate MedGene and to use for the analysis with the micro-array data. Spearman rank-correlation coefficients were used to assess the linear relationship between LPF and micro-array fold change in expression level.

**Global Analysis.** Diseases with at least 50 related genes were selected for clustering analysis, and the LPF scores were normalized with total score for each disease. Hierarchical clustering was done with the "Cluster" software and the clustering result was visualized using "TreeView" (<http://rana.lbl.gov/EisenSoftware.htm>).

**Breast Tissue Micro-Arrays.** Eighty-nine breast cancer samples (79% ER-positive) and 7 normal breast tissue samples were selected from the Harvard Breast SPORC frozen tissue repository and were representative of the spectrum of histological types, grades, and hormone receptor immuno-phenotypes of breast cancer. Biotinylated cRNA, generated from the total RNA extracted from the bulk tumor, was hybridized to Affymetrix U95A oligo-nucleotide micro-arrays. These micro-arrays consist of 12 400 probes, which represent approximately 9000 genes. Raw expression values were obtained using GENE-CHIP software from Affymetrix, and then further analyzed using the DNA-Chip Analyzer (dChip) custom software.

### Results

**Automated Indexing of Medline Records by Disease and Gene.** To study the gene-disease associations in the literature, we first compiled complete lists for human diseases and human genes. To index all Medline records that were relevant to human diseases, the Medical Subject Heading (MeSH) index of Medline records was utilized. MeSH is a controlled medical vocabulary from the National Library of Medicine and consists of a set of terms or subject headings that are arranged in both an alphabetic and an hierarchical structure. Medline records are reviewed manually and MeSH terms are added to each with software assistance.<sup>9,10</sup> Twenty-three human disease category headings along with all of their child terms (see the Supporting Information, Supplemental Table 1, or visit [http://hipseq.med.harvard.edu/MedGene/publication/s\\_Table1.html](http://hipseq.med.harvard.edu/MedGene/publication/s_Table1.html)) were selected from the 2002 MeSH index creating a list of 4033 human diseases.

No index comparable to the MeSH index exists for genes, and thus, it was necessary to apply a string search algorithm for gene names or symbols found in Medline text. A complete list of genes, gene names, gene symbols, and frequently used synonyms were collected from the LocusLink database at NCBI,<sup>11,12</sup> which contains 53 259 independent records keyed by an official gene symbol or name (June 18<sup>th</sup>, 2002). For the purposes of this study, no distinction was made between genes and their gene products. Authors often use the same name for both, differentiating the two only by the use of italics, if at all. For the intended use of this study, this lack of distinction is unlikely to have a large effect and may in fact be beneficial.

Initial attempts to search the literature using these lists revealed several sources of false positives and false negatives (Table 1). False positives primarily arose when the searched term had other meanings, whereas false negatives arose from syntax discrepancies necessitating the development of filters to reduce these errors. The syntax issues were readily handled by including alternate syntax forms in the search terms. The false positive cases, caused by duplicative and unrelated meanings for the terms, were more difficult to manage. Where possible, case sensitive string mapping reduced inappropriate citations. In many cases, however, this was not sufficient and the terms had to be eliminated entirely, thereby reducing the false positive rate but unavoidably under-representing some genes.

For the purposes of data tracking, a primary gene key was selected to represent all synonyms that correspond to each gene. Medline records were indexed with a primary gene key when any synonym for that key was found in the title or abstract. Case-insensitive string mapping was used for all searches except as noted above. No additional weight was



Table 1. Systematic Sources of False Positives and False Negatives in Unfiltered Data\*

source of error	error type	example	filter solution
gene symbol/name is not unique	false positive	MAG—myelin associated glycoprotein MAG—malignancy-associated protein	eliminate this term
gene symbol is unrelated abbreviation	false positive	PA—pallid homologue (mouse), pallidin (also abbrev. for Pennsylvania)	eliminate this term
gene symbol/name has language meaning	false positive	WAS—Wiskott—Aldrich Syndrome (also the word "was")	case-sensitive string search
nonstandard syntax	false negative	BAG-1 instead of BAG1	add dash term
unofficial gene name/symbol	false negative	P53 instead of TP53	add all gene nicknames
nonspecified gene name	false negative	estrogen receptor instead of Estrogen receptor 1	add family stem term

\* In preliminary studies, Medline was searched for co-occurrence of genes and diseases and the resulting output was evaluated to identify error sources that were amenable to global filters. Each error source is categorized by the type of error it causes: false positives are suggested relationships that are not real and false negatives are real relationships that are underrepresented. The filter solutions used are indicated. Note that in some cases, the filter solution itself introduces error. In general, error rates maximized sensitivity, even at the expense of specificity if needed.

added for multiple occurrences of a term or the co-occurrence of multiple synonyms for the same gene key.

Medline records were searched with all qualified gene identifiers, such as the official/preferred gene symbol, the official/preferred gene name, all gene nicknames and all syntax variants. In situations where there are several members of a gene family or splice variants, some authors prefer to use a shortened gene family name, e.g., estrogen receptor instead of estrogen receptor 1 (*ESR1*), creating a source of false negatives. For this reason, gene family stem terms were created for all genes that have an alpha or numerical suffix (e.g., *IL2RA*, *TGFβ*, *ESR1*, etc.) and then used to search the literature. The family stem terms were handled separately from the specific gene names so that it would be clear when linkages were made to the gene family versus a specific member in that family.

To improve performance and accuracy, some pre-selection was applied to the records that were scanned. First, review articles were eliminated to avoid redundant treatment of citations. Second, non-English journals were removed because the natural language filters were only relevant to English publications. Finally, journals unlikely to contain primary data about gene-disease relationships were also removed (e.g., *Int. J. Health Educ.*, *Bedside Nurse*, and *J. Health Econ.*). Together, these filters reduced the 12 198 221 Medline publications (July 2002) by 37%.

**Ranking the Relative Strengths of Gene-Disease Associations.** In total, there were 618 708 gene-disease co-citations, in which 16% (8297) of all studied genes had been associated to a disease and 96% (3875) of all diseases had been associated to at least one gene. To rank the relative strengths of gene disease relationships, we tested several different statistical methods and examined the results. With the exception of the relative risk estimates, the methods provided similar results with respect to the rank order of the gene-disease association strengths. However, after comparing the results to other databases and after consulting disease experts, the log of the product of frequency (LPF) was selected for further analysis because it gave the best results overall.

**Validation of MedGene.** In developing this tool, it was important to minimize the number of missed genes (false negatives) and misclassified genes (false positives). However, in situations when these goals were in conflict, inclusiveness was prioritized. To determine the false negative rate in MedGene, breast cancer was used as a test case because it was associated with more genes than any other human disease and because

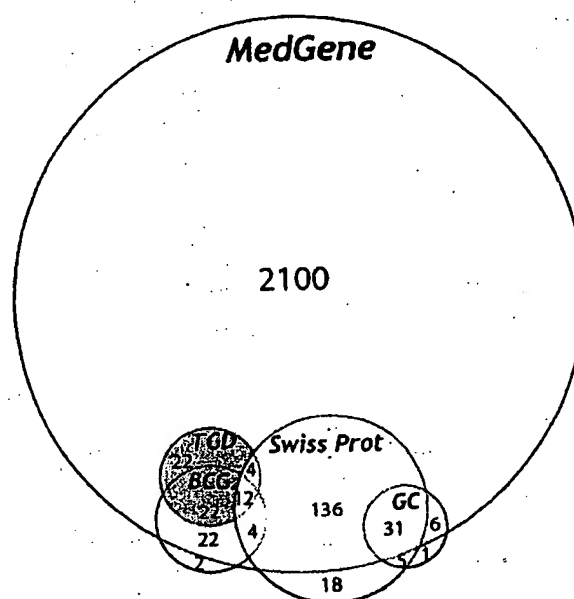


Figure 1. Estimation of the false negative rate by comparison with hand-curated databases. The breast cancer-related genes identified by MedGene were compared with those listed in several other databases including the Tumor Gene Database (TGD),<sup>2</sup> the Breast Cancer Gene Database (BCG),<sup>1</sup> GeneCards (GC)<sup>17</sup> and Swissprot.<sup>18</sup> Genes were considered false negatives if they were represented in at least one of these other databases and not in MedGene and their link to breast cancer was supported by at least one literature reference. All literature references were verified by manual review to confirm their validity. The number of genes in each database or shared by more than one database is indicated. The false negative rate was calculated by genes missed at MedGene (26)/total number of nonoverlapping genes in other databases (285).

there were several public databases that link genes to breast cancer. We compared the list of breast cancer-related genes from MedGene to these databases, illustrated in Figure 1. Among the 285 distinct breast cancer-related genes that were supported by at least one literature citation in these hand-curated databases, 26 were absent from MedGene, suggesting a false negative rate of approximately 9%. To determine why these were missed, all literature references for these genes (80

papers) were reviewed manually (see the Supporting Information, Supplemental Table 2, or visit [http://hipseq.med.harvard.edu/MedGene/publication/s\\_Table\\_2.html](http://hipseq.med.harvard.edu/MedGene/publication/s_Table_2.html)). Among these papers, most false negatives were caused by nonstandard gene terms or gene terms eliminated by our specificity filters. Few genes were missed because they were only mentioned in review papers (0.4%) or they appeared only in the body of the manuscript but not the abstract or title (1.1%). Of note, MedGene identified approximately 2000 additional breast cancer-related genes not listed in any other database.

To assess the false positive error rate, two complementary approaches were used: a detailed analysis of one disease and a global examination of 1000 diseases. The detailed approach examined the false positive error rate and its sources, whereas the global approach tested whether the overall results made biomedical sense.

Using the LPF, 1467 genes related to prostate cancer were assembled in rank order. We then retrieved approximately 300 Medline records each for the highest ranked 100 and the lowest ranked 200 genes and manually reviewed the titles and abstracts to determine the verity of the association. Nearly 80% of the highest ranked 100 genes fell into one of the five categories that reflect meaningful gene-disease relationships (see the Supporting Information, Supplemental Table 3, or visit [http://hipseq.med.harvard.edu/MedGene/publication/s\\_Table\\_3.html](http://hipseq.med.harvard.edu/MedGene/publication/s_Table_3.html)). Among the lowest ranked 200 genes, approximately 70% reflected true relationships. Of the 600 records reviewed, there were only two in which the association between the gene and the disease was described as negative. Both were genes with very low scores. In both cases, the authors did not argue the absence of any relationship, but rather that a particular feature of the gene or protein was not shown to be related to human prostate cancer.<sup>13,14</sup>

The coincidence of some gene symbols with medical abbreviations, chemical abbreviations and biological abbreviations resulted in most of the false positives (see the Supporting Information, Supplemental Table 4, or visit [http://hipseq.med.harvard.edu/MedGene/publication/s\\_Table\\_4.html](http://hipseq.med.harvard.edu/MedGene/publication/s_Table_4.html)), emphasizing the importance of the filters that were added in the search algorithm (Table 1). Without the filters, the false positive rate more than doubled, and the false negative rate rose dramatically (data not shown). For example, among the papers about breast cancer, there were only 12 Medline records that referred to *ESR1* and 10 to *ESR2*, whereas almost 2000 papers mentioned estrogen receptor without specifying *ESR1* or *ESR2*; this latter group was detected by the family term filter.

To further validate these results, a global analysis of the gene-disease relationships described by MedGene was performed. For this experiment, it was reasoned that the more closely related the diseases are to one another, the more they will be related to the same gene sets. Thus, if the relationships defined by MedGene accurately reflected the literature, then an unsupervised hierarchical clustering of the gene data should group diseases in a manner consistent with common medical thinking. Conversely, if the clustered diseases do not make sense biologically or medically, it may reflect excessive false positives, false negatives, or inappropriate scoring of the data.

To execute this experiment, the gene sets and the corresponding LPF values for 1000 randomly selected diseases (each with at least 50 gene relationships) were used as a dataset for clustering the diseases. A review of the results showed that the resulting disease clusters were indeed logical based upon common medical knowledge (see the Supporting Information,

Supplemental Figure 1, or visit [http://hipseq.med.harvard.edu/MedGene/publication/s\\_Figure\\_1.html](http://hipseq.med.harvard.edu/MedGene/publication/s_Figure_1.html)). For example, in one such cluster shown in Figure 2, diabetes and its complications grouped together and were also closely linked to diseases associated with starvation states.

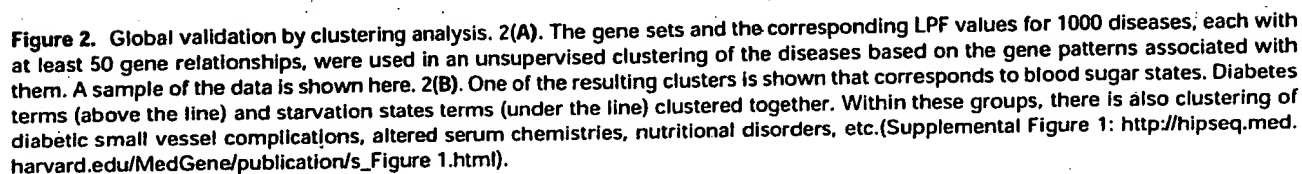
The number of genes associated with a given disease can be estimated by adjusting the MedGene number up by the false negative rate (~9%) and down by the false positive rate (~26% on average). Using this, the average disease has  $103.7 \pm 45.3$  (mean  $\pm$  s.d.) genes associated with it, although the range is quite broad with 2359 genes related to breast cancer, 2122 genes related to lung cancer and no genes related to a number of diseases.

**Applying MedGene to the Analysis of Large Datasets.** Access to a comprehensive summary of the genes linked to human diseases provided an opportunity to analyze data obtained from a high-throughput experiment. We compared the MedGene breast cancer gene list to a gene expression data set generated from a micro-array analysis comparing breast cancer and normal breast tissue samples. Micro-array analysis identified 2286 genes that had greater than a 1-fold difference in mean expression level between breast cancer samples and normal breast samples. Using MedGene, we sorted the 2286 genes into four classes: 555 genes directly linked to breast cancer in the literature by gene term search (first-degree association by gene name); 328 genes directly linked by family term search (first-degree association by family term); 1021 genes linked to breast cancer only through other breast cancer genes (second-degree association); and 505 genes not previously associated with breast cancer. (See the Supporting Information, Supplemental Figure 2, or visit [http://hipseq.med.harvard.edu/MedGene/publication/s\\_Figure\\_2.html](http://hipseq.med.harvard.edu/MedGene/publication/s_Figure_2.html).) Among the 505 previously unrelated genes, 467 were either newly identified genes or genes that had not previously been associated with any disease. Among the remaining 38 genes, 9 had been related to other cancers, specifically esophageal, colon, uterine, skin, and cervix.

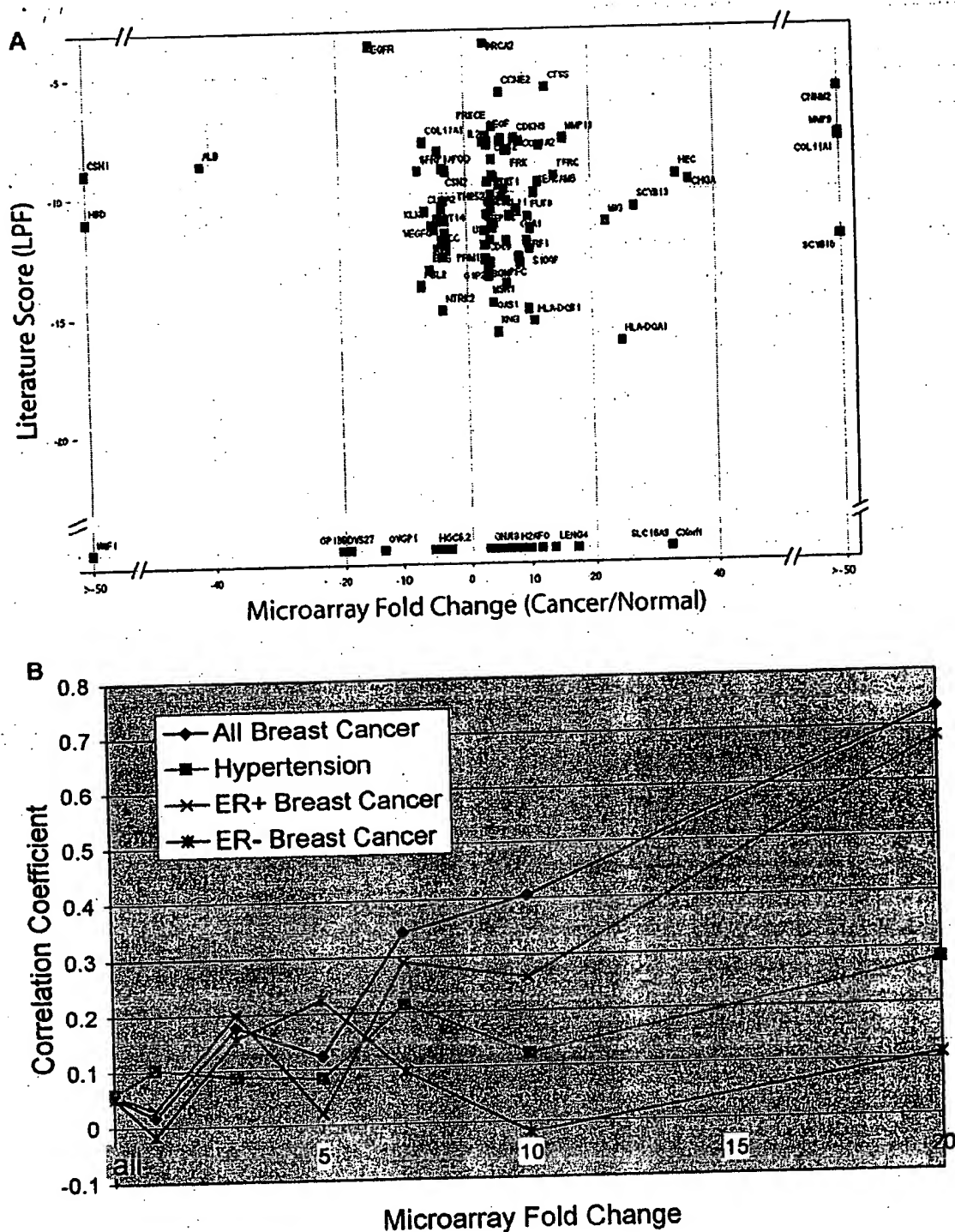
To determine whether the genes highlighted by the micro-array analysis were more likely to have been previously linked to breast cancer in the literature, we created a two-dimensional plot of the fold change of expression level between breast cancer and normal tissue versus the literature score (LPF) (Figure 3A). There was a broad spread of expression changes among the genes directly linked to breast cancer ranging from less than 1-fold change (68%) to over 40-fold (0.3%). Notably, the majority of genes with greater than 10-fold expression changes were linked to breast cancer by first-degree association.

Among all 754 genes directly linked to breast cancer in the literature, there was no correlation between LPF and micro-array fold change ( $r = 0.018$ ,  $p$ -value = 0.62). However, when we stratified the analysis based on the magnitude of the fold change, we observed an increasing trend in correlation (Figure 3B) suggesting that genes with a more substantial change in expression level were more likely to have a stronger association in the literature. For genes that had 10-fold change or more in expression level, the correlation increased to 0.41 ( $p$ -value = 0.05).

When we evaluated the micro-array data separately for ER positive and ER negative tumors, the trend in correlation between fold change and literature score was highly dependent on estrogen receptor status. Interestingly, there was a similar trend in correlation for ER positive tumors, but no trend in correlation for ER negative tumors.



disease unrelated to breast cancer. As expected, we did not observe an increasing trend in correlation for hypertension.



**Figure 3.** Relationship between literature score and functional data for breast cancer. **3A.** The data from an expression analysis of breast samples for breast tumors and normal breast tissue were analyzed to indicate the fold difference of expression level between breast tumor and normal sample (cutoff  $\geq 3$ -fold change). The fold changes were plotted against the literature score for the same gene set. Green dots represent first-degree association by gene search, blue dots represent first-degree association by family search and red dots represent no-association. Some well-studied genes, such as BRCA2 (pink circle), are not reflected by a substantial difference in expression level. Furthermore, the majority of genes that have no association with breast cancer in the literature had less than 10-fold expression changes (shaded area). **3B.** The Spearman rank-correlation coefficients between literature score (LPF) and the fold change of expression level between tumor and normal breast samples (y-axis) in relation to the amount of fold change of expression level (x-axis). Gene rank lists were generated for breast cancer (blue) and hypertension (pink). Correlations were also computed between the breast cancer gene LPF scores and fold change expression data among estrogen receptor positive tumors only (light blue) and estrogen receptor negative tumors only (purple).

Table 2. Top 25 Genes Related to Selected Human Diseases\*

breast neoplasms	hypertension	rheumatoid arthritis	bipolar disorder	atherosclerosis
estrogen receptor	<i>REN</i>	<i>RA</i>	<i>ERDA1</i>	apolipoprotein.
<i>PCR</i>	<i>DBP</i>	<i>TNFRSF10A</i>	<i>SNAP29</i>	<i>APOE</i>
<i>ERBB2</i>	<i>LEP</i>	<i>CRP</i>	<i>PFKL</i>	<i>LDLR</i>
<i>BRCA1</i>	<i>AGT</i>	<i>AS</i>	<i>DRD2</i>	<i>ELN</i>
<i>BRCA2</i>	<i>INS</i>	<i>ESR1</i>	<i>TRH</i>	<i>ARG1</i>
<i>EGFR</i>	kallikrein	<i>HLA-DRB1</i>	<i>IMPA2</i>	<i>APOB</i>
<i>CYP19</i>	<i>ACE</i>	<i>DR1</i>	<i>HTR3A</i>	<i>APOA1</i>
<i>TFF1</i>	endothelin	interleukin	<i>DRD3</i>	<i>MSR1</i>
<i>PSEN2</i>	<i>S100A6</i>	<i>TNF</i>	<i>REM</i>	<i>LPL</i>
<i>TP53</i>	<i>BDK</i>	<i>IL6</i>	<i>KCNN3</i>	<i>PON1</i>
<i>CES3</i>	<i>DIANPH</i>	collagen	<i>DRD4</i>	plasminogen
<i>CEACAM5</i>	<i>SAR1</i>	<i>IL1A</i>	<i>HTR2C</i>	activator inhibitor
<i>ERBB3</i>	<i>PIH</i>	<i>ACR</i>	<i>RELN</i>	<i>PLG</i>
cyclin	<i>CD59</i>	<i>TNFRSF12</i>	<i>DBH</i>	vascular cell
<i>COX5A</i>	<i>ALB</i>	<i>IL2</i>	<i>MAOA</i>	adhesion molecule
cathepsin	<i>CYP11B2</i>	<i>CHI3L1</i>	<i>COMT</i>	<i>ATOH1</i>
<i>ERBB4</i>	<i>MAT2B</i>	<i>IL8</i>	<i>HTR2A</i>	<i>VWF</i>
<i>TRAM</i>	angiotensin receptor	interleukin 1	<i>SYNJ1</i>	<i>INS</i>
<i>CCND1</i>	<i>AGTR2</i>	matrix	<i>INPP1</i>	<i>ARG2</i>
<i>EGF</i>	<i>NPPA</i>	metalloproteinase	<i>NEDD4L</i>	<i>ABCA1</i>
<i>MUC1</i>	<i>LVM</i>	interferon	<i>FRA13C</i>	<i>OLR1</i>
insulin-like	<i>DBH</i>	<i>IL4</i>	transducer of	collagen
<i>BCL2</i>	<i>NPY</i>	<i>IL17</i>	<i>ERBB2</i>	<i>MCP</i>
mucin	<i>POMC</i>	<i>MMP3</i>	<i>BAIAP3</i>	lipoprotein
<i>FGF3</i>	neuropeptide	<i>SIL</i>	<i>ATP1B3</i>	<i>APOA2</i>
			<i>DRD5</i>	intercellular
				adhesion molecule
				<i>RAB27A</i>

\* MedGene results for the top 25 genes associated with breast neoplasms, hypertension, rheumatoid arthritis, bipolar disorder, and atherosclerosis, respectively, ranked by LPF scores. The hyperlink to all the papers co-citing the gene and the disease is available at MedGene website (<http://hipseq.med.harvard.edu/MedGene/>).

## Discussion

The Human Genome Project heralded a new era in biological research where the emphasis on understanding specific pathways has expanded to global studies of genomic organization and biological systems. High-throughput technologies can provide novel insight into comprehensive biological function but also introduces new challenges. The utility of these technologies is limited to the ability to generate, analyze, and interpret large gene lists. MedGene, a relational database derived by mining the information in Medline, was created to address this need. MedGene users can query for a rank-ordered list of human gene-disease relationships (Table 2) for one or more diseases. Each entry is hyperlinked to the original papers supporting each association and to other relevant databases.

MedGene is an innovative extension of previous text mining approaches. Perez-Iratxeta et al. used the GO annotation and their chromosomal locations to predict genes that may contribute to inherited disorders.<sup>8</sup> MedGene takes a broader view and includes all diseases and all possible gene-disease relationships. Furthermore, MedGene utilizes co-citation to indicate a relationship rather than GO annotation, which is limited to the subset of genes that have GO annotation. Our approach is complementary to that taken by Chaussabel and Sher, who used the frequency of co-cited terms to cluster genes into a hierarchy of gene-gene relationships.<sup>6</sup>

A unique aspect of this tool is the ability to assess the relative strengths of gene-disease relationships based on the frequency of both co-citation and single citation. This presupposes that most co-citations describe a positive association, often referred to as publication bias<sup>15</sup> and is supported by our observations

that negative associations are rare (Supplemental Table 3: <http://hipseq.med.harvard.edu/MedGene/publication/s-Table 3.html>). Of course, relationships established by frequency of co-citation do not necessarily represent a true biological link; however, it is strong evidence to support a true relationship.

Another important feature of MedGene is the implementation of software filters that substantially reduced the error rate. We estimate that less than 10% of all associations were missed and at least 70% of even the weakest associations were real. For this study, all of the filters that we applied were general ones, e.g., expanding the list of all gene names to address the different syntax forms used by different journals, eliminating gene names that correspond to common English words, etc. The majority of the remaining search term ambiguities were idiosyncratic and difficult to identify systematically without causing a significant rise in false negatives. Alternative approaches, such as the examination of the nearest neighbor terms, need to be considered to further reduce the false positive rate.

It is not uncommon to see expression changes in microarray experiments as small as 2-fold reported in the literature. Even when these expression changes are statistically significant, it is not always clear if they are biologically meaningful. When comparing expression levels of disease to normal tissue, one expects an enrichment of known disease-related genes to appear in the altered expression group. MedGene provided a unique opportunity to test this notion in the context of existing knowledge on a novel breast cancer micro-array dataset. For genes displaying a 5-fold change or less in tumors compared to normal, there was no evidence of a correlation between altered gene expression and a known role in the disease. This

## research articles

Table 3. Genes with Large Expression Changes in ER- but Not in ER+ Breast Tumors

gene symbol	fold change (ER+)	fold change (ER-)
KRTHB1	1.0	610.8
BR53	1.2	89.4
DKK1	1.2	69.8
ZIC1	1.9	59.6
TLR1	1.0	38.5
KIAA0680	2.6	33.2
CDKN3	1.0	30.6
EBI2	4.0	27.9
GZMB	3.8	21.9
STK18	4.7	18.6
GPR49	1.0	14.6
MYO10	1.6	14.4
LAD1	-1.0	13.5
POLE2	4.2	13.0
HMG4	4.4	12.9
BCL2L11	-1.2	12.3
LRP8	2.9	12.2
CCNB2	1.0	11.8
CCNE2	4.0	11.6
FGB	-4.3	11.1
KNLS6	2.9	10.9
H1F5	3.0	10.2
SERPINH2	4.6	10.2
YAP1	1.0	10.0
LPHB	-1.3	-10.4
TCEA2	-1.1	-10.8
TFF1	1.3	-11.4
COL17A1	-4.1	-15.7
POP5	1.1	-16.2
BPAG1	-4.6	-22.3
PDZK1	-1.1	-36.8
VEGFC	-2.8	-51.5
MUC6	-1.4	-64.9
SERPINA5	-1.0	-83.1
MEIS1	-1.6	-85.9
CA12	2.4	-150.3

Table 3. MedGene identified a set of relatively understudied, yet highly expressed genes in ER negative, but not ER positive breast tumors. All of these genes have either never been co-cited with breast cancer or have a weak association except those marked with an \*.

reflects the many genes whose role in breast cancer may not involve large changes in expression in sporadic tumors (e.g., *BRCA1* and *BRCA2*) and genes whose modest changes in expression may be unrelated to the disease. Strikingly, among genes with a 10-fold change or more in expression level, there was a strong and significant correlation between expression level and a published role in the disease, providing the first global validation of the micro-array approach to identifying disease-specific genes.

The results derived from MedGene have two implications. First, a careful hunt for corroborating evidence of a role in breast cancer should precede any further study of genes with less than 5-fold expression level changes. Second, any genes with 10-fold changes or more are likely to be related to breast cancer and warrant attention. It is likely that this threshold will change depending on the disease as well as the experiment.

Interestingly, the observed correlation was only found among ER-positive tumors, not ER-negative. This may reflect a bias in the literature to study the more prevalent type of tumor in the population. Furthermore, this emphasizes that caution must be taken when interpreting experiments that may contain subpopulations that behave very differently. The MedGene approach identified a set of relatively understudied, yet highly expressed genes in ER-negative tumors that are worthy of further examination (Table 3).

In conclusion, we have developed an automated method of summarizing and organizing the vast biomedical literature. To our knowledge, the resulting database is the most comprehensive and accurate of its kind. By generating a score that reflects the strength of the association, it provides an important tool for the rapid and flexible analysis of large datasets from various high-throughput screening experiments. Furthermore, it can be used for selecting subsets of genes for functional studies, for building disease-specific arrays, for looking at genes common to multiple diseases and various other high-throughput applications. In the future, it will be possible to enhance the utility of the MedGene database by building links between genes and other MeSH terms as well as other biological processes and concepts, such as cell division and responses to small molecules.

**Acknowledgment.** We would like to thank P. Braun, L. Garraway, J. Pearlberg, and other members of our institute for helpful discussion. Many thanks to the NLM (National Library of Medicine) for licensing of MEDLINE and the annotation effort of adding MeSH indexes for MEDLINE abstracts. This work was funded by grants from the Breast Cancer Research Foundation and an NHLBI PGA Grant (Vol HL65582-02).

**Supporting Information Available:** Twenty-three human disease category headings along with all of their child terms selected from the 2002 MeSH index (Supplemental Table 1); analysis of the causes of false negatives in MedGene (Supplemental Table); meaningful gene-disease relationships found in MedGene (Supplemental Table 3); causes for incorrect assignment of gene indexes (Supplemental Table 4); a review of the results, showing that the resulting disease clusters were indeed logical (Supplemental Figure 1); and a review of the results showing that among the 505 previously unrelated genes, 467 were either newly identified genes or genes that had not previously been associated with any disease (Supplemental Figure 2). This material is available free of charge via the Internet at <http://pubs.acs.org> and at the web sites mentioned in the text.

## References

- (1) Baasiri, R. A.; Glasser, S. R.; Steffen, D. L.; Wheeler, D. A. *Oncogene* 1999, 18, 7958-7965.
- (2) Steffen, D. L.; Levine, A. E.; Yarus, S.; Baasiri, R. A.; Wheeler, D. A. *Bioinformatics* 2000, 16, 639-649.
- (3) Marcotte, E. M.; Xenarios, I.; Eisenberg, D. *Bioinformatics* 2001, 17, 359-363.
- (4) Ono, T.; Hishigaki, H.; Tanigami, A.; Takagi, T. *Bioinformatics* 2001, 17, 155-161.
- (5) Jenssen, T. K.; Laegreid, A.; Komorowski, J.; Hovig, E. *Nat. Genet.* 2001, 28, 21-28.
- (6) Chaussabel, D.; Sher, A. *Genome Biol.* 2002, 3, RESEARCH0055.
- (7) Gibbons, F. D.; Roth, F. P. *Genome Res.* 2002, 12, 1574-1581.
- (8) Perez-Iratxeta, C.; Bork, P.; Andrade, M. A. *Nat. Genet.* 2002, 31, 316-319.
- (9) Funk, M. E.; Reid, C. A. *Bull. Med. Libr. Assoc.* 1983, 71, 176-183.
- (10) Humphrey, S. M.; Miller, N. E. *J. Am. Soc. Inf. Sci.* 1987, 38, 184-196.
- (11) Maglott, D. R.; Katz, K. S.; Sicotte, H.; Pruitt, K. D. *Nucleic Acids Res.* 2000, 28, 126-128.
- (12) Pruitt, K. D.; Maglott, D. R. *Nucleic Acids Res.* 2001, 29, 137-140.
- (13) Wadelius, M.; Andersson, A. O.; Johansson, J. E.; Wadelius, C.; Rane, E. *Pharmacogenetics* 1999, 9, 333-340.
- (14) Adam, R. M.; Borer, J. G.; Williams, J.; Eastham, J. A.; Loughlin, K. R.; Freeman, M. R. *Endocrinology* 1999, 140, 5866-5875.
- (15) Montori, V. M.; Smieja, M.; Guyatt, G. H. *Mayo Clin. Proc.* 2000, 75, 1284-1288.
- (16) Denenberg, V. H. *Statistics Experimental Design for Behavioral and Biological Researchers*; Wiley-Liss: New York, 1976.
- (17) Rebhan, M.; Chalifa-Caspi, P.; Prilusky, J.; Lancet, D. *Trends Genet.* 1997, 13, 163.
- (18) Baloch, A.; Apweiler, R. *Nucleic Acids Res.* 2000, 28, 45-48. PR0340227



## Genomic and proteomic analysis of the myeloid differentiation program

Zhang Lian, Le Wang, Shigeru Yamaga, Wesley Bonds, Y. Beazer-Barclay, Yuval Kluger, Mark Gerstein, Peter E. Newburger, Nancy Berliner, and Sherman M. Weissman

Although the mature neutrophil is one of the better characterized mammalian cell types, the mechanisms of myeloid differentiation are incompletely understood at the molecular level. A mouse promyelocytic cell line (MPRO), derived from murine bone marrow cells and arrested developmentally by a dominant-negative retinoic acid receptor, morphologically differentiates to mature neutrophils in the presence of 10  $\mu$ M retinoic acid. An exten-

sive catalog was prepared of the gene expression changes that occur during morphologic maturation. To do this, 3'-end differential display, oligonucleotide chip array hybridization, and 2-dimensional protein electrophoresis were used. A large number of genes whose mRNA levels are modulated during differentiation of MPRO cells were identified. The results suggest the involvement of several transcription regulatory factors not

previously implicated in this process, but they also emphasize the importance of events other than the production of new transcription factors. Furthermore, gene expression patterns were compared at the level of mRNA and protein, and the correlation between 2 parameters was studied. (Blood. 2001;98:513-524)

© 2001 by The American Society of Hematology

### Introduction

Studies of normal myeloid maturation from many laboratories have identified genes that may play critical roles in myeloid differentiation.<sup>1-4</sup> Current studies suggest that these events are dependent on a cascade of molecular changes that involve complex modulation of mRNA transcription. Furthermore, studies of acute leukemia have suggested that the disease arises from the accumulation of myeloid precursors arrested at early stages of differentiation and associated, in many cases, with chromosomal rearrangements that alter the structure of specific transcription factors.<sup>5</sup> Nevertheless, the molecular events underlying the production of mature myeloid cells are not well understood and appear to use interacting pathways and networks, the elucidation of which requires an extensive description of the molecular components available to the myeloid cell.

An extensive body of information is accumulating with respect to gene expression profiles of mammalian cells. However, much of the information available in public databases has been accumulated by the use of techniques such as single oligonucleotide chips or cDNA arrays that measure fewer than 6000 of potentially 30 000 to 120 000 transcripts. The more limited range of analyses reported by the serial analysis of gene expression (SAGE)<sup>6,7</sup> technique accurately estimates changes in levels of the more abundant mRNAs but requires extensive redundant analyses to measure changes in the patterns of expression of scarce mRNAs. We have used a modified polymerase chain reaction (PCR)-based cDNA differential display (DD) method in which single restriction fragments derived from the 3' end of cDNAs are separated on a sequencing gel.<sup>8,9</sup> Bands from the gel can be identified initially by sequencing, but then

comparison of patterns from different samples can be made without further sequencing. This sensitive and reproducible method detects, in principle, most cDNAs regardless of whether they are represented in existing databases.

Systematic analysis of the function of genes can also be performed at the protein level. This approach has the advantage of being closest to function, because proteins perform most of the reactions necessary for the cell. The most common method of proteome analysis is the combination of 2-dimensional gel electrophoresis (2DE) to separate and visualize protein and mass spectrometry (MS) for protein identification.<sup>10</sup> Several such analyses of yeast and of normal or malignant mammalian cells have been performed. To date, however, there have been few studies in which both mRNA and protein have been compared by applying analyses to the same samples. The studies of Anderson<sup>11</sup> and Gygi<sup>12</sup> showed that there is not a good correlation between mRNA and protein levels, in yeast or human liver cells. However, other analyses disagree with this conclusion (Greenbaum et al, manuscript submitted, and Fletcher et al<sup>14</sup>). Furthermore, global correlations between changes in mRNA and protein levels have not been examined during the execution of any developmental program.

The MPRO cell line was derived by transduction of a dominant-negative retinoic acid receptor construct into normal mouse bone marrow cells. It is a granulocyte-macrophage colony-stimulating factor (GM-CSF)-dependent line arrested at a promyelocytic stage of development.<sup>15,16</sup> After treatment with all-*trans* retinoic acid (ATRA) most of the cells acquire the morphology of mature

From the Department of Genetics, Boyer Center for Molecular Medicine, the Section of Hematology, Department of Internal Medicine, and the Department of Molecular Biophysics and Biochemistry, Yale University School of Medicine, New Haven, CT; the Department of Pediatrics, University of Massachusetts Medical School, Worcester, MA; and Gene Logic, Gaithersburg, MD.

Submitted December 4, 2001; accepted March 28, 2001.

Supported by grants from the National Institutes of Health (NIH) (CA42559) and Gene Logic (A143558, DK54369, and HL63357). Z.L. is supported by NIH grant HL 63357. P.E.N. is supported by NIH grant DK 54369, grants from the Arthritis Foundation and the Charles H. Hood Foundation, and the Pierce Family Cancer Research Fund. M.G. is supported by the Kock Foundation and

by NIH grant GM54160-04.

L.W. and S.Y. contributed equally to this research.

Reprints: Sherman M. Weissman, Department of Genetics, Boyer Center for Molecular Medicine, Yale University School of Medicine, Rm 336, 295 Congress Ave, New Haven, CT 06536-0812; e-mail: sherman.weissman@yale.edu.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 U.S.C. section 1734.

© 2001 by The American Society of Hematology

neutrophils and begin to produce neutrophil lactoferrin and gelatinase, 2 proteins characteristic of neutrophil secondary granules.<sup>17</sup> As such, it offers a valuable model for studying neutrophil differentiation in vitro.

We now report the analysis of mRNA expression changes during the process of MPRO cell maturation to neutrophils and compare the results with a limited analysis of cellular protein composition. mRNA expression changes were studied by combining the use of oligonucleotide arrays and DD. A database (dbMC) with comprehensive genomic information for myeloid differentiation program was constructed (accessible at <http://www.bioinfo.mbb.yale.edu/expression/neutrophil>). We have grouped the changes in mRNA levels of a large number of genes into 6 patterns, with implications for the genetic program of myeloid differentiation.

We also compared 2-dimensional high-resolution gel electrophoretograms from control cells and cells differentiated for 72 hours in the presence of ATRA. Fifty protein spots whose relative intensity changed prominently during differentiation were examined by mass spectrometry. The results suggest a poor correlation between mRNA expression and protein abundance, indicating that it may be difficult to extrapolate directly from individual mRNA changes to corresponding ones in protein levels (as estimated from 2DE).

## Materials and methods

### Cell lines

MPRO cells and HM-5 cells provided by Dr Schickwahn Tsai (Fred Hutchinson Cancer Research Center, Seattle, WA)<sup>15</sup> were used throughout the study. The cells proliferated continuously as a GM-CSF-dependent cell line at 37°C in Iscoves modified Dulbecco medium (Gibco BRL, Grand Island, NY) supplemented with 5% to 10% fetal calf serum (Gibco BRL) and 10% HM-5-conditioned medium as a source of GM-CSF. Morphologic differentiation of the blocked MPRO promyelocytes was induced by treatment with 10  $\mu$ M ATRA (Sigma, St Louis, MO). Controls were cultured in the absence of ATRA but with the same volume of vehicle (ethanol).

### RNA isolation and differential display

After exposure to 10  $\mu$ M ATRA for 0, 24, 48, or 72 hours, total cellular RNA was isolated from MPRO cells using TRIzol reagent (Life Technologies, Gaithersburg, MD). cDNA was then synthesized using a T-7 Sal-Oligo d(T) 32 primer as described previously.<sup>8,18</sup> The double-stranded cDNA was digested with 1 of 9 different restriction enzymes (*ApaI*, *BglII*, *BamHI*, *EagI*, *EcoRI*, *HindIII*, *XbaI*, *KpnI*, and *SphI*) and ligated to Y-shaped adaptors with a complementary overhang. DNA fragments were then amplified by PCR as described previously.<sup>8,18</sup> PCR products were separated on a sequencing gel of 6% polyacrylamide with 7 M urea. The gel was dried and exposed to x-ray film. Genes from differential display gels, whose maximum intensity changes equaled 2+ on a scale of 1+ to 8+, were recorded as significantly changed.<sup>19</sup> Individual DNA bands were recovered from the gels, amplified by PCR, and sequenced.

### Oligonucleotide chip analysis of RNA samples

Ten micrograms total RNA from each sample (0, 24, 48, or 72 hours) was used to prepare cDNA. This cDNA was transcribed with T7 RNA polymerase to prepare a fluorescently labeled probe.<sup>20,21</sup> Each sample was hybridized to mouse array chip (MultiK Array; Affymetrix, Santa Clara, CA) containing oligonucleotide probe sets corresponding to approximately 7000 known genes or ESTs represented by UniGene clusters.<sup>22</sup> cDNAs were considered present if their probe set results were rated as such by the GeneChip software (Affymetrix) and if the average difference (AD) between perfect match and mismatch probe pairs was not less 100 U. If a

gene was represented by more than one array probe set, the average of all probe sets for the gene was taken. Genes with AD values between 100 and 200 were considered unchanged because of their low expression levels. Those genes with AD values equal to or more than 200 U at one time point were further studied by rescaling, threshold, and normalization methods described in the MIT Center for Genome Research Web site.<sup>13</sup> A value of 20 was assigned to any gene with an AD below 20 at some time point.

### Bioinformatics and database development

All the sequences or gene fragments were searched using Blast against GenBank and TIGR gene indices. A database of genes or ESTs whose expression levels changed during myeloid differentiation was constructed containing information for each band or gene. This included GenBank matches, Locus Link or Unigene clusters, expression patterns, tissue distribution, synonym(s) protein name, gene name(s), notations of possible functions, poly A signal and sequence quality, and hyperlinks to the database searches, sequence trace files, and related references. All gene data were then gathered into a cluster file. Supplementary information is available at <http://bioinfo.mbb.yale.edu/expression/neutrophil>.

### Classification and analysis of DNA fragments

Sequences from differential display analyses were classified as representing known genes, ESTs, genomic sequences, or novel genes as described.<sup>19,23</sup> Known genes from both differential display and arrays were clustered into 27 functional categories and searched against SWISS-PROT (<http://www.expasy.ch/nrc.ca/cgi-bin/sprot-search-full>) or PIR (<http://www.pir.georgetown.edu/>). Information such as function, subcellular location, family and superfamily classification, map position, similarity, synonym(s) protein name, gene name(s), and so on was recorded in a variety of databases.

### Northern blot analysis

Thirty micrograms total cellular RNA per lane from time-course MPRO cells were loaded onto 1.2% formaldehyde-agarose gels, then transferred to Hybond-N+ membranes (Amersham Pharmacia Biotech, Uppsala, Sweden). After standard prehybridization, membranes were hybridized overnight at 65°C with radiolabeled cDNA probes (ordered from Research Genetics according to their dbEST Image ID). Membranes were washed at a final stringency of 60°C in 0.1  $\times$  SSC.

### Immobilized pH gradient 2-dimensional gel electrophoresis and mass spectrometry

Induced MPRO cells collected at 0 and 72 hours were lysed with lysis buffer (540 mg urea, 20 mg dithiothreitol, 20  $\mu$ L Pharylytic [3-10], 1.4 mg phenylmethylsulfonyl fluoride, 1  $\mu$ g each aprotinin, leupeptin, pepstatin A, and antipain 50  $\mu$ g TLCK, and 100  $\mu$ g TPCK/1 mL). We applied 100  $\mu$ L each MPRO cell lysate (2.5  $\times$  10<sup>6</sup> cells/100  $\mu$ L) to immobilized pH gradient (IPG) strips (pH 3-10 L; Amersham Pharmacia Biotech), and IPG electrophoresis was conducted for 16 hours (20 100 Vb) using an Immobiline Drystrip Kit (Amersham Pharmacia Biotech). Electrophoresis in the second dimension was carried out in a 12% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gel with the Laemmli-SDS continuous system in a Proteom II xi 2-D cell (Bio-Rad) run at 40 mA constant current for 4.5 hours. Proteins were detected by Brilliant Blue G-250 staining.<sup>24</sup> Protein spots were excised from the gel and digested with trypsin. ACTH clip (average [M+H]<sup>+</sup> 2466.70) and bradykinin (average [M+H]<sup>+</sup> 1061.23) were used for calibration of peptide masses. One microliter sample digest was mixed with 1.0  $\mu$ L  $\alpha$ -cyano-4-hydroxy cinnamic acid (4.5 mg/mL in 50% CH<sub>3</sub>CN, 0.05% TFA) matrix solution and 1  $\mu$ L calibrants (100 fmol) each. The spectra of the peptides were acquired in reflector/delayed extraction mode on a Voyager-DE STR mass spectrometer (PerSeptive Biosystems, Foster City, CA). Peptides were identified using the ProFound search engine.<sup>19</sup>



## Results

### Differentiation of MPRO cells

Figure 1 illustrates the morphologic changes in an MPRO cell population representative of those used for RNA expression analysis. Undifferentiated MPRO cells resembled promyelocytes under the light microscope (Figure 1A). After induction with ATRA for 24 hours, the cells morphologically differentiated into metamyelocytes (Figure 1B). At 48 hours, the cells further developed into metamyelocytes and band neutrophils (Figure 1C). At 72 hours, nearly 100% of MPRO cells became mature neutrophils (Figure 1D).

### Identification of mRNAs by differential display assay

MPRO cellular mRNA was analyzed at 0, 24, 48, and 72 hours after ATRA treatment. Nine restriction enzymes were used in a 3'-end DD approach. During MPRO differentiation, 1109 fragments corresponding to 837 transcripts were found to change substantially in expression levels (Figure 2). These represented approximately 279 known genes, 112 ESTs, and 59 putative new genes, each with a perfect or fair polyadenylation signal at an appropriate distance from the oligo-dT priming site. The gene information detected by DD was collected in database dbMCD.

### Identification of mRNAs by oligonucleotide chip assay

We used an oligonucleotide chip containing 13 179 probe sets corresponding to approximately 7000 murine genes to analyze patterns of mRNA expression in the same RNA samples used for DD. The information obtained by oligonucleotide arrays was collected in the database dbMCA.

We clustered the genes by their similarity to idealized expression patterns. For instance, the expression pattern of an ideal gene that is overexpressed (high) at time 0 and underexpressed (low) at 24, 48, and 72 hours, would be high-low-low-low (HLLL). Overall we have  $(2^4 - 2)$  idealized patterns excluding HHHH and LLLL. Pearson correlation was used as the

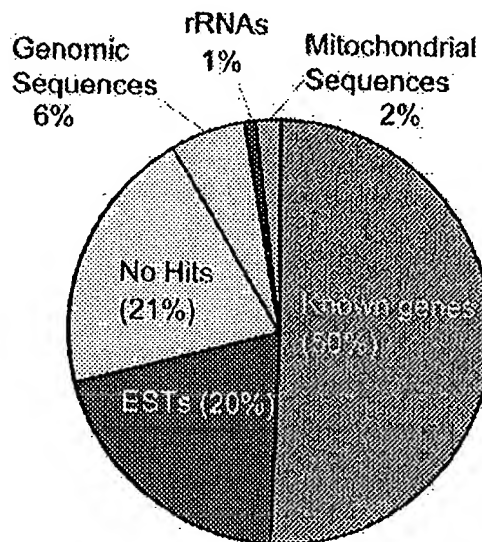


Figure 2. Distribution of genes obtained by DD assay. MPRO cell mRNA was analyzed at 0, 24, 48, and 72 hours after ATRA treatment; 1109 fragments corresponding to 837 transcripts were found to change substantially in expression levels. The total 937 transcripts were classified into 6 categories according to the bioinformatic analysis. Percentages show the gene distributions in these 6 categories. Information for each transcript was collected in database dbMCD.

measure of similarity of each gene expression pattern,  $x = (x_1, x_2, x_3, x_4)$  to each of the 14 idealized patterns  $y = (y_1, y_2, y_3, y_4)$ . The 4 entries of  $x$  and  $y$  corresponded to the 4-dimensional gene expression levels at 0, 24, 48, and 72 hours, respectively. Each gene was assigned to a cluster labeled by the idealized pattern that had the maximal correlation with that gene. We selected only genes that hybridized well compared with the background (considered "present" by GeneChip software) and had maximal AD amplitude greater than 200 U in at least 1 of the 4 stages. We further tabulated the 14 patterns according to whether the gene expression changed at early (0-hour), intermediate (24- and 48-hour), and late (72-hour) time points and whether gene expression monotonically increased (up-regulated), monotonically decreased (down-regulated), or was not monotonic (transient). Table 1 shows 8 clusters of 104 genes that had significant changes of mRNA levels, arranged according to the temporal stage and the monotonic/transient changes of expression levels.

Principal component analysis determined whether we could comprehensively present multidimensional data (4-dimensional in our case) in a simple 2-dimensional graph. First, we found the 4 principal components, which were the axes of the most compact 4-dimensional ellipsoid that encompassed the 4-dimensional cloud of data. Each axis was a different linear combination of the original 4 variables. Then we verified that the first 2 principal components (the first 2 largest axes of the ellipsoid) captured most (95.2%) of the variation of the data. Therefore, the data could be faithfully projected (with a minor loss of information) into a 2-dimensional graph, with the 2 largest principal components as the x- and y-axes. As shown in Figure 3, genes tend to coalesce in clusters, according to their labels determined by their similarity to an ideal expression pattern. In summary, a genomic (global) picture of the distribution of genes according to their similarity to predetermined idealized multidimensional expression patterns is concisely displayed in a 2-dimensional graph.

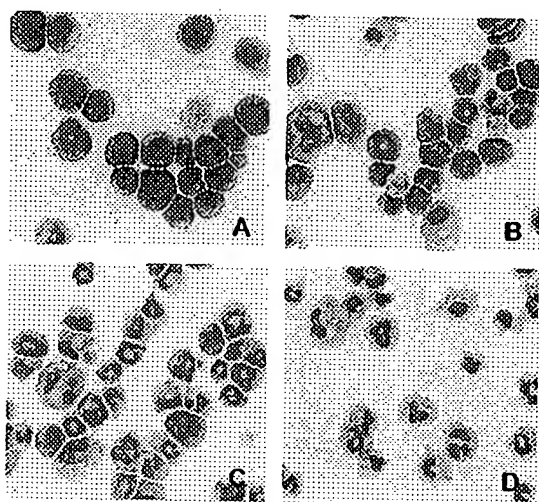


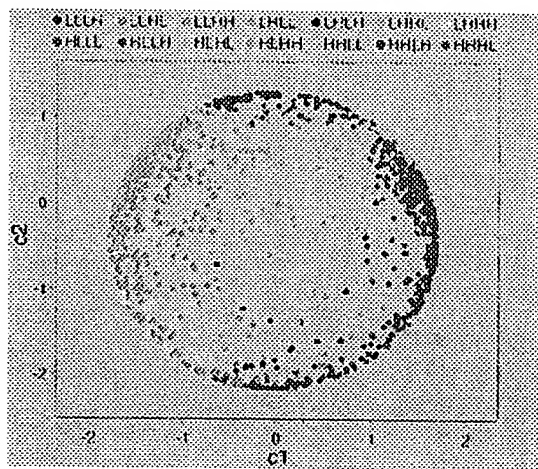
Figure 1. Morphology of MPRO cells during differentiation. MPRO cells were induced as described in "Materials and methods," concentrated by cytopsin, and Wright-Giemsa stained. (A) Uninduced MPRO cells. (B) MPRO cells induced with ATRA for 24 hours. (C) MPRO cells induced with ATRA for 48 hours. (D) MPRO cells induced with ATRA for 72 hours.

**Table 1. Genes differently regulated during the different stages of mouse promyelocytic cell line differentiation process**

Category	Timing		
	Early	Middle	Late
Up-regulation	LHHH (n = 16) <i>Mad P2rx1 Rgb2 Il1r2 Lcn2 Ilpr5</i> <i>Cebpb H2-D Etch16 Zyx</i>	LLHH (n = 6) <i>Pral Cybb Plc Pira5 Cd53 Ilmgr2</i>	LLH (n = 13) <i>Il1a Csf1r li Ctsl S100a8 L-CCR Ctss</i> <i>Aldo1 Rac2 Fpr1 Ctss Ubb Ptmb4</i>
Down-regulation	HLLL (n = 11) <i>Tcrg-V4 Ly64 Ctsg Spi2-1 Mcpt8</i> <i>Myc Myb Tlr4 Npm1 Erh Hsp60</i>	HHLL (n = 1) <i>Mpo</i>	HHHL (n = 37) <i>Actx Ir2 EL2 Rpl19 Actb Ly6a Arf1 Hist2</i> <i>Psm2a Gnas Zfp36 Il4ra Ltbr Shldg1</i> <i>Max Rps8 Csf2rb1 Sipi Tdrex1 Tpi Btf3</i> <i>Cnfr Gys3 Slc10a1 Clsb Sepp1 Rtn3</i> <i>Ccrnb2 S100a9 Cf11 Hist5-2ax Rela</i> <i>Copa Gstm1 Gnb2-ret1 Grn RPL8</i>
Transient		LLHL (n = 9) <i>Sei1 Klr2 Pira6 Plrb Lst1 Ltf Sema4d Stat6 Mmp9</i> LHHL (n = 17) <i>Cebpa Lyzs Fcgr3 Arf5 Lamp1 Stat3 Csf2ra Osl</i> <i>Actg Sfpi1 Gpx3 Plprc Prtn3 Irf1 Rps6ka1</i> <i>Ltb4r Myln</i>	

Arrays of Affymetrix Mu11k containing 13103 probe sets corresponding to 12002 GenBank accessions were used for hybridization. Arrays were hybridized with streptavidin-phycoerythrin (Molecular Probes) biotin-labeled RNA and scanned. Intensity for each feature of the array was captured using Genechip software (Affymetrix), and a single raw expression level for each gene was derived from the 20 probe pairs representing each gene using a trimmed mean algorithm. For each gene, an AD of 24-, 48-, and 72-hour samples was calibrated by dividing the slopes of the linear regression line for a graph with the x-axis the AD of 0-hour probe sets and the y-axis the AD of the respective time point (24, 46, or 72 hours). A threshold of 20 U was assigned to any gene with a calculated expression level below 20 because discrimination of expression below this level could not be performed with confidence.<sup>36</sup> Each gene expression profile was categorized as described in Tables 3, 4, and 5. For the 4 time points, the minimum AD of the relatively higher group (MIN-H) was divided by the maximum AD of the relatively low group (MAX-L), and those genes whose MIN-H/MAX-L greater than 2 were selected as meaningfully regulated. Genes were sorted in descending order based on the MIN-H/MAX-L. Genes in boldface are those whose expression level was in the top 20% (ie, maximum AD of 4 time points greater than 3000), and genes in italics are those in the bottom 20% (ie, maximum AD of 4 time points less than 300). The differentiation period was grouped into 3 stages: early (0-hour), middle (24-hour and 48-hour), and late (72-hour) stages.

AD indicates average difference; gene symbols are expanded in an Appendix at the end of this article.

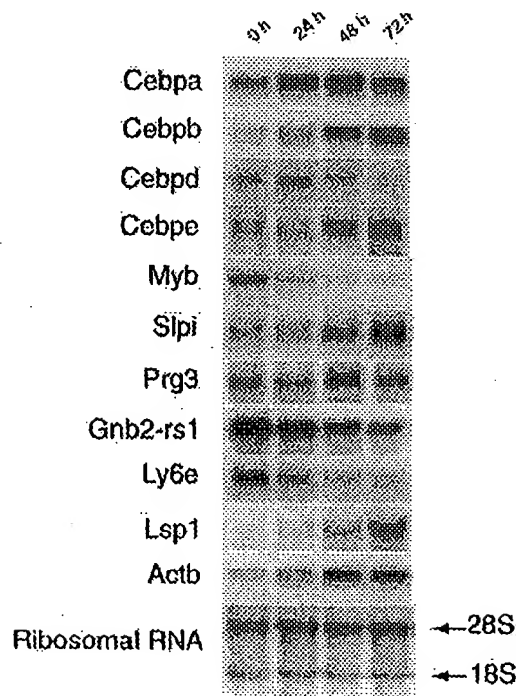


**Figure 3. Gene clusters in the first 2 principal component spaces.** Principal component analysis allowed us to present the multidimensional data (in this case, 4-dimensional data of each gene expression pattern) in a simple 2-dimensional graph. We derived the 4 principal components, which are a linear combination of the standardized expression intensities (zero mean and unit variance) at 0, 24, 48, and 72 hours. The first 2 principal components captured most of the variation of the data (approximately 85%). Therefore, the data can be displayed (with a minor loss of information) in a 2-dimensional graph. The first and second principal components,  $c_1$  and  $c_2$ , are given by the linear combinations  $c_1 = 0.747 \cdot n_1 - 0.11 \cdot n_2 - 0.658 \cdot n_3 + 0 \cdot n_4$  and  $c_2 = 0.278 \cdot n_1 + 0.353 \cdot n_2 + 0.233 \cdot n_3 - 0.863 \cdot n_4$ , where  $n_1, n_2, n_3$ , and  $n_4$  are the rescaled and standardized expression levels at 0, 24, 48, and 72 hours, respectively. The axes legends  $c_1$  and  $c_2$  stand for the first 2 principal components. In this paper we used the Pearson correlation to measure the similarity of each gene with the idealized expression patterns, as opposed to the Euclidean distance we used in a previous work,<sup>19</sup> because clusters were better separated using this measure. In both cases, we presented the data in the 2-dimensional space of the lowest principal components. The data had a tendency to be circularly distributed when we used the Pearson correlation as a distance measure.

We have previously demonstrated a correlation coefficient of 0.93 between visual estimates of changes in band intensity on DD and PhosphorImager System (Molecular Dynamics, Sunnyvale, CA) estimates of band intensity and a correlation coefficient of 0.88 between hybridization intensity changes of mRNA on Northern blot analyses and changes in band intensity on DD.<sup>19</sup> In a few cases there were clear discrepancies in the pattern of expression of a gene, as estimated by DD and by oligonucleotide chip analysis. We chose the 6 most extreme cases and examined the levels of mRNA change for these genes by Northern blot analysis (Figure 4). In 5 cases, the Northern blot results agreed with the results of the DD analysis, whereas the results of Gnb2-rs1 disagreed with the oligonucleotide array but duplicate bands from DD showed a relatively high level of expression in the 0 time sample that did not correlate with the Northern blot (Table 2). One possible explanation for these findings was the change in the relative use of different polyadenylation sites after the addition of ATRA to the MPRO cells.

### Constructing a database for mRNA level changes during myeloid differentiation

Based on the data obtained above, an in-house database (dbMC) was constructed that included 2 subdatabases, dbMCd and dbMCe, for collecting gene information from DD or oligonucleotide arrays, respectively. Each entry in dbMC is accompanied by a so-called executive summary. The linkage between dbMCd and dbMCe was established by UniGene ID and cluster ID. dbMC contains the temporal expression patterns of genes during the MPRO cell differentiation process, including not only products represented in public databases but also novel transcripts.



**Figure 4.** Northern blot analysis of selected mRNAs. Equivalent amounts of RNA from MPRO cells induced by ATRA at different time points (0 hour, 24 hours, 48 hours, and 72 hours) were resolved by formaldehyde-agarose gel electrophoresis, stained to verify the amount of loading. Eleven genes were separately probed on the RNA filters. The gene symbol of each probe was listed at the left of a related Northern blot result. Detailed information on these 11 probes was listed in Table 5. One of the RNA-blotted membrane photographs is shown with methylene blue-stained 28S and 18S RNA subunits demonstrating the quality and quantity of RNA loaded in individual lanes.

#### Analysis of gene expression patterns during MPRO differentiation

Many of the genes identified in this study were found in myeloid cells or were implicated in myeloid development for the first time. We detected 8 cytokines<sup>25</sup> and chemokines whose mRNA levels changed more than 5-fold by arrays and 2-fold by DD during the maturation of MPRO cells (see our Web site, <http://bioinfo.mbb>.

[yale.edu/expression/neutrophil](http://yale.edu/expression/neutrophil)). Among these were 2 members of the CC chemokine family. Interleukin-1 $\alpha$  (IL-1 $\alpha$ ) was up-regulated at the late stage of differentiation (LLLH pattern, Table 1).

mRNA for approximately 52 receptors was detected by one or the other method. A number of the receptors known to be present on mature neutrophils showed late induction of mRNA, and their levels of induction were high, indicating that the expression of these products is a prominent event late in neutrophil maturation (Table 3). Rarely was mRNA for receptors down-regulated, consistent with myeloid maturation being accompanied by increasing responsiveness of the cell to a variety of external stimuli.

#### Expression of mRNA for granule proteins

Neutrophils contain several types of granules that develop at different stages of myeloid maturation.<sup>3,17,26</sup> Levels of mRNAs encoding secondary granule proteins, such as lactoferrin, increased as the cells matured (Table 4). The level of mRNA for Mmp9, reported as a tertiary granule protein, increased markedly between 24 and 48 hours after the induction of differentiation, whereas mRNAs for secondary granule proteins either increased less markedly or showed a maximum increase by 24 hours. mRNAs for several primary granule constituents, such as myeloperoxidase and cathepsin G, were present in unstimulated cells and decreased as the cells matured. There was a discrepancy in the measurements of proteoglycan mRNA by DD and oligonucleotide chips, but Northern blots showed that it reached a peak at 48 hours and then declined (Figure 4). Cathepsin D is reported as a primary granule protein, but its pattern of mRNA expression more closely resembled that of secondary granule constituents. In addition to known granule components, mRNAs for several other cathepsins were up-regulated during myeloid differentiation, in parallel with or later than the tertiary granule protein mRNAs.

#### mRNAs for transcription factors

Transcription factor genes, including several identified at the sites of consistent chromosome rearrangements in acute myeloid leukemia, have been implicated in normal myeloid differentiation and in the expression of neutrophil proteins.<sup>2,5,27</sup> However comprehensive information concerning the expression of these transcription factors during myeloid development is not readily available. Therefore, we compared gene names and identifiers in our databases to those of the transcription factor database Transfac (<http://>

**Table 2.** Expression patterns of genes detected by Northern blot analysis

Gene symbol	Gene accession	AD value by array				Intensity by DD			
		0 h	24 h	48 h	72 h	0 h	24 h	48 h	72 h
Cebpa	M62362	33	212	182	44	—	—	—	—
Cebpb	X62600	390	1248	1380	1993	—	—	—	—
Cebp d	X61800	157	262	168	430	—	—	—	—
Cebp e	—	—	—	—	—	—	—	—	—
Myb	M12848	882	356	230	435	—	—	—	—
Sipi	U73004	617	501	783	402	1	2	3	3
Prg3	W45834	153	259	339	345	5	1	1	2
Gnb2-rs1	X75313	4231	3623	3215	3403	4	4	1	1
Ly6e	U04268	3061	5381	2844	1262	3	2	1	1
Lsp1	M90316	65	376	840	28	2	3	5	6
Actb	X03765	3095	3588	3976	2434	1	2	3	2

Gene symbol and gene accession refer to National Center for Biotechnology Information databases and, in particular, to Locus Link. AD value is the average difference in the value of hybridization intensity between the set of perfectly matched oligonucleotides and the set of mismatched oligonucleotide in the oligonucleotide array. Band intensities from DD were semiquantified on a scale from 1 (+) to 8 (+++++). These estimates are shown as boldface numbers in this table.<sup>19</sup> Both AD value and intensity of genes were studied at 4 time points corresponding to MPRO cells induced for the indicated times.

DD indicates differential display; MPRO, mouse promyelocytic cell line; for gene symbols, see the Appendix at the end of this article.

Table 3. Receptors expressed during myeloid differentiation process

Maximal fold change	Gene symbol	Gene accession	AD value by array			
			0 h	24 h	48 h	72 h
Less than 2	Bzrp	D21207	641	658	881	887
	Cmkar4	X99581	508	447	378	604
	Cry	M34173	433	384	506	506
	Csf2rb1	M34397	318	345	410	241
	Hlr5a	Z18278	168	272	273	339
	M6pr	X64068	536	409	408	649
	MPPR	AA116789	232	84	63	381
	TCRGB	M28053	165	212	244	299
	Trfslfa	M59377	0	1	1	1
2 or more, less than 3	Cmkbr1	U28494	221	244	504	638
	Ctrr	X72305	121	200	256	355
	Csf2ra	M85078	171	372	402	254
	Ebl3	AF013114	187	270	428	148
	Grid1	D10171	128	164	150	257
	Ifngr	J05265	141	263	327	251
	Il2rg	U21795	205	184	231	477
	Ldr	X64414	1399	1653	1665	3968
	P40-8	J02870	849	677	381	640
	Plaur	X62701	312	443	476	734
	Rarg	M34476	102	113	114	218
	Srb1	U37799	126	232	132	258
	Cr2	M29281	83	138	243	77
	Csf2rb2	M29855	209	249	437	111
3 or more, less than 4	Fcgr1g	J05020	2398	2766	3365	8751
	Fcgr2b	X04648	1703	1652	1431	4605
	Ifngr2	U69599	1	2	2	3
4 or more, less than 5	Nr4a1	X16995	96	188	202	401
5 or more	Il1r2	X59769	482	1796	2672	3818
	C5r1	L05630	185	434	808	1078
	Drd2	X55674	0	0	0	219
	Fcgr3	M14215	1	1	1	2
	Fpr1	L22161	0	69	141	671
	GCR	AA240711	2	0	0	0
	L-CCR	AA034646	48	175	314	2056
	NMDARGB	AAB20211	2	2	0	0
	P2rx1	X84896	79	346	530	744
	Pira1	U96682	0	43	172	378
	Pira5	U96666	274	391	954	1874
	Pira6	U96687	122	635	2014	1716
	Pib	U96689	191	445	966	747
	Sell	M25324	46	104	570	20
	Tcrg-V4	M54998	1650	78	65	315

Receptors are identified as present whose maximal AD values were more than or equal to 250 U in this study. Genes were sorted by their expression patterns as follows: first by the average difference value, then by the difference between minimum and maximum AD for the 4 time points, and last by the alphabetical order of gene symbols. Genes were ordered according to the maximal fold change of AD values. Abbreviations of gene names are taken from gene symbols listed in the Locus Link portion of the National Center for Biotechnology Information database where available. Numbers in bold denote those gene expression patterns obtained by differential display rather than by oligonucleotide array assays. The other information is presented as in the legend to Table 2.

AD indicates average difference; gene symbols are expanded in an Appendix at the end of this article.

www.transfac.gbf-braunschweig.de/TRANSFAC) and determined which factors contained in this database were present at detectable levels in MPRO cell mRNA, using Affymetrix software for the criteria for inclusion of mRNAs from approximately 200 murine transcription factors probe sets on the oligonucleotide chip. Of these, 54 were expressed and 13 showed changes of 3-fold or more in chip signal (Table 5).

The changes in certain transcription factors, such as the moderate down-regulation of *myb* and *myc* and the up-regulation of the Max dimerization protein MAD, were consistent with the shift of the cells

from a proliferative to a differentiated state.<sup>28</sup> Some changes are more difficult to explain, such as the up-regulation of DP1, a partner for E2f factors in the regulation of S-phase genes, and the mild up-regulation of the *Id* genes, commonly associated with an inhibition of differentiation by competition with bHLH transcriptional activators.<sup>29</sup>

The C/EBP family has been extensively studied with respect to myeloid differentiation.<sup>2,30</sup> Absolute levels of the C/EBP  $\alpha$  and  $\delta$  mRNAs were low, probably at the borderline of significance for the oligonucleotide chip assay, whereas the level of C/EBP  $\beta$  appeared higher. In addition, there were discrepancies between the chip

Table 4. Granule constituents expressed during mouse promyelocytic cell line cell differentiation

Granule constituent	Gene symbol	Gene accession	AD value by array			
			0 h	24 h	48 h	72 h
Azurophil (primary) granules						
	Man2c1	AA161869	178	134	99	164
	Ctsb	M65270	442	480	595	389
	Ctsd	X52886	214	1087	1828	2784
	Cteg	M96801	1509	405	46	285
	E12	U04962	658	1273	843	157
	E1a2	AA589016	47	159	134	163
	Gus-s	M63836	544	226	266	254
	Lyzs	M21050	0	1	1	3
	Mcp18	X78545	831	268	66	491
	Mpo	X15378	3788	3009	776	692
	Prp	X16133	2621	2653	2920	9659
Possible granule proteins						
	Ctsc	AA144887	252	194	342	576
	Ctse	X97399	1	3	4	5
	Ctsh	U06119	45	124	195	156
	Ctsl	X06086	16	11	31	237
	Ctss	AA089333	12	9	88	463
Specific secondary granules						
	Cpa3	J05118	621	270	90	601
	Cd36l2	AB008553	113	93	157	187
	Cnlp	X94353	80	479	704	626
	Cybb	U43384	8	24	81	128
	Ea2	---	0	1	1	2
	Fpr1	L22181	178	220	235	845
	Itgb2	X14951	0	2	4	2
	Lcn2	W13166	916	3513	3831	6036
	Ltf	J03298	19	162	333	138
	MBP	W45834	5	1	1	2
	Mmp13	X66473	44	43	72	178
	Ngp	L37297	2661	4782	2311	6912
Tertiary granules						
	Mmp9	Z27231	0	1	2	2

Shown are the possible granule protein cDNAs represented on the oligonucleotide arrays, sorted by their expression patterns as follows: first by the average difference AD value, then by the granule types, and last by the alphabetical order of gene symbols. Data are presented as described in the legend to Table 3.

AD indicates average difference; gene symbols are expanded in an Appendix at the end of this article.

estimates and the mRNA levels observed by Northern blotting with specific probes for these genes. In particular, the latter method, more sensitive and specific, showed that C/EBP  $\alpha$  began to decline in the most mature cells, whereas C/EBP  $\delta$  mRNA declined progressively beginning at 24 hours after the onset of differentiation.

C/EBP  $\epsilon$  is a more recently cloned C/EBP family member. Previous studies indicated it is expressed in a large array of human leukemia cell lines blocked at various stages of differentiation and that it is up-regulated during granulocytic differentiation.<sup>31</sup> A C/EBP  $\epsilon$  probe was not included in the oligonucleotide chips, and this mRNA was not detected by DD. Therefore, we examined the C/EBP  $\epsilon$  expression patterns by quantitative PCR and Northern blot analysis (Figure 4). C/EBP  $\epsilon$  exon 1 was PCR amplified from MPRO RNAs using primers RY48 (AGCCCCGACACCCTTGATGA) and RY49 (TGCCACACT-GCGGGCAGACAG).<sup>32</sup> The results showed that C/EBP  $\epsilon$  is expressed throughout myeloid differentiation, with expression levels increased moderately in the later stages.

We detected a number of other transcription factors that are broadly expressed or that have been reported in other studies of hematopoiesis (Table 5). Some of the factors that were most strongly induced during differentiation have been studied in other contexts but not previously implicated in hematopoiesis, such as a mammalian homologue to the *Drosophila* enhancer of split gene, a transcriptional silencer. The mammalian gene is expressed at relatively high levels as measured by the oligonucleotide chip and

is a candidate for mediation of the silencing of growth-related genes in the maturing neutrophil. Another candidate transcriptional silencer, Tif1b, may serve as a corepressor for the KRAB domain family of zinc finger transcription factors and also may mediate binding of the heterochromatin protein HP1 to DNA.<sup>33</sup>

There were 26 transcription factors whose mRNAs showed no significant changes by oligonucleotide chip analysis and were not identified as differentially regulated genes by differential display assays. PU.1, a factor necessary for the production of neutrophils and the expression of several neutrophil genes,<sup>34</sup> showed less than a 3-fold increase in mRNA, below the threshold for a significant change. Other candidate hematopoietic transcription factors, such as PEBP1aB2 (AML1), GATA-1, and SP-2, were represented on the oligonucleotide chips, but their mRNA levels were so low that they were reported as absent in this study. The possibility that small changes in the levels or ratios of some transcription factors could produce marked changes in transcription potentially limits the ability of data generated by present methods to explain transcriptional changes during differentiation.

#### Protein expression patterns of MPRO cells during ATRA induction

We visually compared the 2DE patterns from MPRO cells at the same time points used for mRNA analysis. In most cases the

Table 5. Transcription modulators presented during myeloid differentiation

Maximal fold change	Gene symbol	Gene accession	AD value by array			
			0 h	24 h	48 h	72 h
Less than 2-fold						
	Zfp11-6	AB020542	2630	2989	2795	2515
	Btf3	W13502	3	3	2	1
	Gata2	AB000096	562	770	472	730
	Hmg1	J04179	337	348	177	232
	Idb1	M31885	455	787	721	637
	Max	M63903	256	224	312	172
	Nfatc2	AA560093	2313	3218	2396	2542
	Pm1	U33626	173	261	329	306
	Rarg	M34476	102	113	114	218
	Rela	M61909	297	260	304	244
	Sox15	W53527	419	461	484	837
	Ybx1	M62857	643	469	472	496
	Zfp162	Y12838	671	734	720	992
2 or more, less than 3						
	Cebpd	X61800	157	262	168	430
	Idb2	M69293	244	210	310	604
	Jund1	W29356	1274	2002	1434	3085
	Lyl1	X57687	399	342	347	891
	Nfe2	L09600	458	743	1042	505
	Nfkb1	L28117	953	2044	1876	2034
	Pbx1	AF020186	611	303	345	212
	stpl1	A34683	375	784	991	529
	Tif1b	U67303	673	659	420	863
	Trp53	P10361	259	149	125	361
	Usf2	U12283	120	185	285	182
	Ybx3	L35549	96	169	210	119
	Zfp216	AA510137	82	151	204	106
3 or more, less than 4						
	Irf1	M21065	85	207	278	188
	Klf2	U25096	62	86	246	77
	Myb	M12848	692	356	230	435
	Stat3	AA396029	484	1057	1912	290
	Tfdp1	Q08639	307	560	505	1093
4 or more, less than 5						
	Cebpb	X62600	390	1248	1360	1903
	Stra14	Y07836	223	383	510	935
5 or more						
	Cebpa	M62362	33	212	182	44
	Grg	X73359	99	565	916	1005
	Mad	X83106	0	111	167	327
	Myc	L00039	314	112	62	173
	Etsi6	W89667	169	386	313	1003
	TBX1	AA542220	0	0	1	2

Shown are the transcription factors identified as present by the oligonucleotide array analysis whose maximal AD between perfect match and mismatch oligonucleotide sets was greater than or equal to 250 U in this study. Data are presented as described in the legend to Table 3.

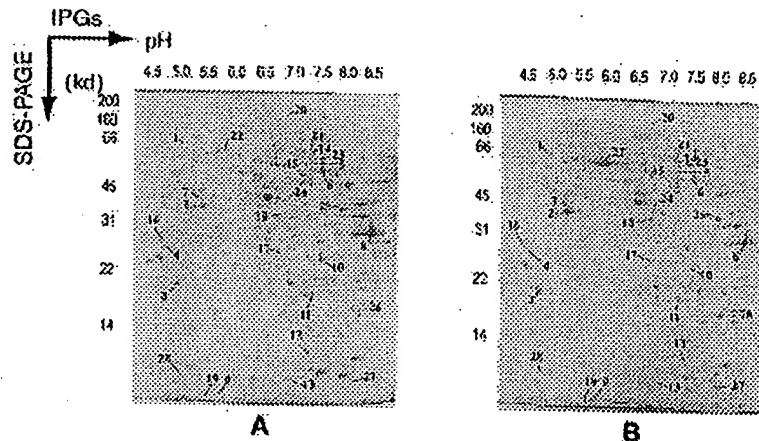
AD Indicates average difference; gene symbols are expanded in an Appendix at the end of this article.

peptides identified for a given protein were derived from regions along the entire length of the protein, indicating the observed products were not the result of proteolytic degradation. These data must be considered with several caveats: membrane and other hydrophobic proteins and very basic proteins are not well displayed by the standard 2DE approach, and proteins present at low levels will be missed.<sup>35</sup> In addition, to simplify MS analysis, we used a Coomassie dye stain rather than silver to visualize proteins, and this decreased the sensitivity of detection of minor proteins. The MS method we used was sufficiently sensitive to identify proteins that could barely be visualized by colloidal blue staining. However, a limitation of the method for the mouse is that the current database lacks predicted amino acid sequences for a substantial fraction of murine genes. In addition, very small proteins give only a few peptides, making statistically confident identification difficult.

Figure 5 shows the analytical colloidal blue-stained 2DE IPG reference maps of differentiated MPRO cells. Expression patterns of more than 500 protein spots were detected and observed through the entire series of gels. Protein spots could easily be cross-matched to each other, indicating the reproducibility of the method. As marked on the gel pictures (Figure 5), 50 proteins with a wide range of molecular weights (1 to 200 kD), isoelectric points (4 to 9), and abundances were subjected to MS protein identification. The results are presented in Table 6.

Comparing the theoretical value of the molecular weight and *pI* of each protein to that of the observed value, we confidently identified 28 proteins in the expected position on the gels (spots 1 to 28). Some of the other proteins with strong matches to the murine databases migrated to a somewhat unexpected *pI* position. Nine spots gave clear peptide peaks on mass spectroscopy but did not match any known gene. Their identification will require amino acid

**Figure 5.** 2DE electrophoretograms of MPRO cells. MPRO cell lysate ( $2.5 \times 10^6$  cell/sample) was loaded for 2DE analysis. Gels were stained with brilliant blue G-collidal dye. (A) 2DE map of uninduced MPRO cell (0 hour). (B) 2DE map of matured MPRO cells (72 hours). Protein spots marked in the maps were considered differentially expressed and were subjected to MS analysis. The resultant protein information is listed in Table 6.



sequence analysis or availability of more extensive murine databases. We searched for the expression patterns of the genes cognate to the expressed proteins in dbMC (Table 6). Nineteen genes were found in dbMC, the mRNA for 5 genes was reported as absent, and 13 genes were present during MPRO differentiation. Comparison of the expression patterns showed only 4 genes of 18 present on the oligonucleotide chips whose expression was consistent at the RNA level and protein level. None of these was on the list of the genes

that were differentially expressed significantly (5-fold or greater change by array or 2-fold or greater change by DD).

## Discussion

We explored the temporal patterns of gene expression during myeloid development. A database has been developed to provide a

**Table 6.** Correlation of expression patterns between mRNA level and protein level

Spot	Protein definition	Gi number	Predicted value		Percentage (%)	2DE pattern		cDNA expression pattern		Ag
			kd	pI		0 h	72 h	0 h	72 h	
1	GRP 78	2505545	72.4	5.1		1	3	1321	1043.3	N
2	Actin, gamma, cytoplasmic	6752954	41.77	5.3	40	3	6	0	2	Y
3	RHO GDI 2	2494703	22.83	4.9	33	3	3	341	441.6	Y
4	Proliferating cell nuclear antigen	7242171	28.77	4.7	42	1	0	544	430.9	Y
5	APS kinase	4038346	69.8	7.1	24	2	1	43	50.7	N
6	Pyruvate kinase 3	6755074	57.9	7.2	48	6	4	3047	5880.3	N
7	Melanoma X-actin	6671509	41.72	5.3	39	1	3	2539	341.3	N
8	Glyceraldehyde-3-phosphate dehydrogenase	6679937	35.79	8.7	39	8	7	3073	5742.3	N
9	Stein 3	461911	10.99	5.9	46	0	4	N/A	N/A	—
10	Guanine nucleotide binding protein, beta-2, related sequence 1	6680047	35.06	7.9	21	4	2	139	303.1	N
11	Triosephosphate isomerase	6678413	28.69	6.9	26	3	3	3312	2650.1	Y
12	Testis-derived c-abl protein	1196524	17.19	7	51	2	3	152	126.9	N
13	RNA binding motif protein 3	7949121	16.59	6.8	25	1	0	626	812.4	N
14	Colapsin response mediator	6681019	62.16	5.4	36	2	0	Absent	Absent	N
15	Lamin A	220474	47.52	6.6	35	2	0	Absent	Absent	N
16	47-kd keratin	52783	35.82	4.8	29	3	0	Absent	Absent	N
17	sid478p	5931555	31.3	6.7	30	1	2	Absent	Absent	N
18	MHC class II H2-IA-beta-5	3168662	28.6	7.1	39	1	2	N/A	N/A	—
19	Androgen-binding protein: subunit alpha	739346	8.04	6.4	68	0	2	Absent	Absent	N
20	Neuronal apoptosis inhibitory protein	5932010	158.7	6	17	1	0	N/A	N/A	—
21	PAD type IV	6755019	74.46	7.2	21	1	3	N/A	N/A	—
22	Human serum albumin homologue	3212625	66.45	5.7	24	0	6	N/A	N/A	—
23	syncip	6576815	62.53	7.2	33	2	1	N/A	N/A	—
24	Transamidinase	1730203	48.22	7.2	31	3	1	N/A	N/A	—
25	PGK crgr phosphoglycerate	1730519	44.54	8.3	47	5	4	1088	1402.3	N
26	Proliferation-associated gene A	6754976	22.16	8.6	53	3	1	N/A	N/A	—
27	Putative peroxisomal antioxidant enzyme	3913065	17	7.8	55	0	3	N/A	N/A	—
28	IgE chain C2 region	2137430	12.1	5.2	38	0	1	N/A	N/A	—

The proteins listed here are represented by the spots marked in the electrophoretograms shown in Figure 5.

Protein definition, Gi number, and predicted value refer to the protein name, accession number, and properties derived from the National Center for Biotechnology Information protein database. The column labeled % shows the percentage of peptides predicted from the protein sequence that were detected by mass spectroscopy. The expression level of protein spots expressed in mouse promyelocytic cell line cell induced by all-trans retinoic acid for 0 hours and 72 hours (Figure 5) were scored on a scale of 1 (+) to 8 (+++++) in the 2DE pattern column. The cDNA expression patterns of the cognate mRNAs are listed in the cDNA expression pattern column abstracted from the dbMC database. The genes not represented on the oligonucleotide arrays were marked as N/A. Ag showed the correlation of gene patterns at mRNA level or protein level. Y indicates agreement and N discrepancy between changes in cDNA and protein spot intensity. The numbers in bold were obtained with DD. 2DE indicates 2-dimensional gel electrophoresis; IgE, immunoglobulin E; DD, differential display.

reference for later research on the molecular mechanisms underlying normal myeloid development.

The MPRO cell system morphologically mimics normal myeloid differentiation and biochemically proceeds further toward mature neutrophils than most other *in vitro* systems. Because the arrest in differentiation of MPRO cells growing in the absence of ATRA is not physiologic, there is a theoretical risk that gene expression in these cells is not coordinated in the way that it is in normal differentiation. It is encouraging that, for the most part, the timing of expression of genes for proteins of the various neutrophil granules is consistent with the timing of the morphologic and biochemical appearance of these granule components during normal myeloid differentiation.

The DD technique provides certain advantages for detecting and comparing mRNA levels in different samples. First, the method is, in principle, similar to competitive RT-PCR, and, with the use of stringent PCR conditions, is expected to be about as reliable. Second, display patterns are reproducible. Third, the method detects the levels not only of RNAs already represented in the database but also of unknown RNA species that may represent "new" genes. Fourth, closely related genes can be distinguished regardless of cross-hybridization, provided there are some single nucleotide differences in the 3' end sequence. Limitations associated with this technique are that numerous gels are necessary to get complete information and that comparison of the levels of different mRNAs is only approximate because of the differential amplification of bands of different size or sequence.

Oligonucleotide chip analysis is a fast and effective means of accessing mRNA expression patterns.<sup>20</sup> Cluster analysis of groups of samples by this approach is effective. However, the present results indicate that alternative methods of verification are desirable before the data on an unexpected change in a particular gene are definitively accepted.

To obtain the broadest range of information from the myeloid differentiation process, both differential display and oligonucleotide chip techniques were applied in the current study. As a result, 65.3% of the observed changes in mRNA levels came from the differential display method and 41.5% came from oligonucleotide chip assays.

Our data showed in general that changes in expression pattern by the 2 methods agreed qualitatively but that there was some quantitative variation. Our results indicate that DD may be a more accurate way to detect changes in levels of gene expression than the oligonucleotide chip assay. However, improvements in the types of oligonucleotides used in arrays may close this gap in the future.

The mRNAs for a limited number of transcription factors vary in a pattern correlating with that of the mRNAs for primary or secondary granule proteins. However, more detailed information is needed, and the underlying mechanisms of granule gene regulation remain unclear. The number of potential positive and negative regulatory factors found here is sufficiently small as to make it feasible to perform *in vivo* studies, such as chromatin immunoprecipitation.

The oligonucleotide chip used in this study focused on known genes, whereas the DD method samples all polyadenylated transcripts. The latter method generated a large number of products not associated with known genes, in part because the mouse genome is not as well represented in the database as the human genome. However, our experience with DD and human mRNAs indicates that substantial fractions of the products represented as ESTs or not represented at all in the public databases are cDNA copies from introns, hnRNA, or other RNA with internal A runs.

Approximately 59 sequences obtained from gel-display bands had significant changes in the level of expression and a sequence that did not match that for any named gene in the public databases.

Of these, 38 had plausible or excellent polyA signals. This is only an approximate estimate of the number of new genes found<sup>35</sup> because a fraction of the mRNAs for known genes still had poor polyA signals. In addition, the full 3' untranslated region is often not known for characterized genes, and in some cases these new genes may prove to be identical to products identified by the oligonucleotide chips when more complete sequences are obtained. At the least, their presence indicates that a substantial fraction of the regulatory or functional circuitry of maturing myeloid cells remains unexplored and that valuable tools for their investigation will emerge from a combination of RNA expression studies and analysis of emerging genomic sequences.

The desired end point for the description of gene expression in a biologic system is not only the analysis of mRNA transcript levels but also the accurate measurement of protein abundance. The developments in 2DE and new MS instrumentation make it possible to accomplish this work rapidly and efficiently. In this study, we attempted to identify a number of the proteins differentially expressed between uninduced and ATRA-differentiated MPRO cells and to examine the relation between mRNA and protein expression levels for these genes representing the same state.

For protein levels based on estimated intensity of Coomassie dye staining in 2DE, there was poor correlation between changes in mRNA levels and estimated protein levels. Other groups have studied the correlation between mRNA and protein levels in yeast and liver cells.<sup>11,12,14</sup> In the liver cell experiments,<sup>11,12</sup> correlation coefficients of 0.4 to less than 0.5 were observed. In an extensive study in yeast,<sup>11,12</sup> the correlation coefficient was high if the most abundant mRNAs and proteins were considered. If a handful of these products was omitted, the remaining correlation coefficient was 0.4 or less. However, one could restore some of the correlation by averaging individual data points into broad proteomic categories.<sup>37</sup>

The discrepancies between mRNA and protein levels in MPRO cells appear to be substantially larger than those observed for yeast. Possible causes for the discrepancies include translational regulation, differential expression of certain mRNAs at various stages of cell growth *in vitro*, post-translational protein modification that varies with the stage of maturation of the cells, and selective degradation or excretion of proteins *in vivo*. Furthermore, here we are focusing on a developmental time-course, whereas the yeast study concentrated on the organism in vegetative growth. New techniques, equipment, and bioinformatic analysis tools must be developed to make such systematic, global, and quantitative analyses feasible.

The initial studies of protein expression presented here provide a cautionary note for efforts to interpret cell composition and function in relation to mRNA levels. Discrepancies we observed between gene expression and protein abundance suggest that selective post-transcriptional controls may be at least as important as changes in mRNA levels in determining the protein composition of neutrophils and that they are phenomena less well explored than transcriptional control. Analysis of mRNA expression patterns is itself only a small beginning toward a genome-wide description of cellular components.

## Acknowledgments

We thank Dr S. Tsai (Fred Hutchinson Cancer Research Center) for his kind gift of the MPRO cell line, Dr Fuki M. Hisanaga (Yale University School of Medicine) for helpful advice, and the staff at Gene Logic Inc for data and support.



## References

- Lawson ND, Berliner N. Neutrophil maturation and the role of retinoic acid. *Exp Hematol*. 1999; 27:1355-1367.
- Tenen DG, Hromas R, Licht JD, Zhang DE. Transcription factors, normal myeloid development, and leukemia. *Blood*. 1997;90:489-519.
- Sigurdsson F, Khanna-Gupta A, Lawson N, Berliner N. Control of late neutrophil-specific gene expression: insights into regulation of myeloid differentiation. *Semin Hematol*. 1997;34:303-310.
- Lenny N, Westendorp JJ, Hiebert SW. Transcriptional regulation during myelopoiesis. *Mol Biol Rep*. 1997;24:157-168.
- Yunis JJ, Tonzer J. Molecular mechanisms of hematologic malignancies. *Crit Rev Oncog*. 1993;4: 161-190.
- Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. *Science*. 1995;270:484-487.
- Stolberg J, Urschiltz J, Urban Z, Boyd CD. A quantitative evaluation of SAGE. *Genome Res*. 2000;10:1241-1248.
- Subrahmanyam YV, Baskaran N, Newburger PE, Weissman SM. A modified method for the display of 3'-end restriction fragments of cDNAs: molecular profiling of gene expression in neutrophils. *Methods Enzymol*. 1999;303:272-297.
- Subrahmanyam YVBK, Yamaga S, Newburger PE, Weissman SM. A modified approach for the efficient display of 3' end restriction fragments of cDNAs. In: Leslie RA, Robertson HA, eds. *Differential Display: A Practical Approach*. Practical Approach Series. Oxford, UK: Oxford University Press. 2000;101-123.
- Appella E, Arnot D, Sakaguchi K, Wirth PJ. Proteome mapping by two-dimensional polyacrylamide gel electrophoresis in combination with mass spectrometric protein sequence analysis. *EXS*. 2000;88:1-27.
- Anderson NL, Anderson NG. Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis*. 1998;19:1853-1861.
- Gygi SP, Rochon Y, Franz BR, Aebersold R. Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol*. 1999;19:1720-1730.
- The Whitehead Institute for Biomedical Research/MIT Center for Genome Research. Molecular Pattern Recognition Web site. Available at: [www.genome.wi.mit.edu/MPR/](http://www.genome.wi.mit.edu/MPR/) analysis.htm#RS. Accessed May 4, 2001.
- Futcher B, Latter GI, Monardo P, McLaughlin CS, Gerbasi J. A sampling of the yeast proteome. *Mol Cell Biol*. 1999;19:7357-7369.
- Tsai S, Collins SJ. A dominant negative retinoic acid receptor blocks neutrophil differentiation at the promyelocyte stage. *Proc Natl Acad Sci U S A*. 1993;90:7153-7157.
- Johnson M, Calazzo T, Molina JM, Donahue R, Gropman J. Inhibition of bone marrow myelopoiesis and erythropoiesis in vitro by anti-retroviral nucleoside derivatives. *Br J Haematol*. 1988;70:137-141.
- Lawson ND, Krause DS, Berliner N. Normal neutrophil differentiation and secondary granule gene expression in the EML and MPRO cell lines. *Exp Hematol*. 1998;26:1178-1185.
- Prashar Y, Weissman SM. Analysis of differential gene expression by display of 3' end restriction fragments of cDNAs. *Proc Natl Acad Sci U S A*. 1995;93:659-663.
- Subrahmanyam YVBK, Yamaga S, Prashar Y, et al. RNA expression patterns change dramatically in human neutrophils exposed to bacteria. *Blood*. 2001;97:2455-2468.
- Chee M, Yang R, Hubbell E, et al. Accessing genetic information with high-density DNA arrays. *Science*. 1996;274:610-614.
- Lipshutz RJ, Chee M, Hubbell E, et al. Using oligonucleotide probe arrays to access genetic diversity. *Biotechniques*. 1995;19:442-447.
- Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ. High-density synthetic oligonucleotide arrays. *Nat Genet*. 1999;21:20-24.
- Mao M, Fu G, Wu JS, et al. Identification of genes expressed in human CD34(+) hematopoietic stem/progenitor cells by expressed sequence tags and efficient full-length cDNA cloning. *Proc Natl Acad Sci U S A*. 1998;95:8175-8180.
- Neuhoff V, Arold N, Taube D, Ehrhardt W. Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis*. 1988;9:255-262.
- Wang Q, Miyakawa Y, Fox N, Kaushensky K. Interferon-alpha directly represses megakaryopoiesis by inhibiting thrombopoietin-induced signaling through induction of SOCS-1. *Blood*. 2000;96:2093-2099.
- Gullberg U, Bengtsson N, Bulow E, et al. Processing and targeting of granule proteins in human neutrophils. *J Immunol Methods*. 1999;232:201-210.
- Nichols J, Nimer SD. Transcription factors, translocations, and leukemia. *Blood*. 1992;80:2953-2963.
- Amali B, Land H. Myc-Max-Mad: a transcription factor network controlling cell cycle progression, differentiation and death. *Curr Opin Genet Dev*. 1994;4:102-108.
- Pagliuca A, Gallo P, De Luca P, Lania L. Class A helix-loop-helix proteins are positive regulators of several cyclin-dependent kinase inhibitors' promoter activity and negatively affect cell growth. *Cancer Res*. 2000;60:1376-1382.
- Yamanaka R, Lekstrom-Himes J, Barlow C, Wynshaw-Boris A, Xanthopoulos KG. CCAAT/enhancer binding proteins are critical components of the transcriptional regulation of hematopoiesis (review). *Int J Mol Med*. 1998;1:213-221.
- Morosetti R, Park DJ, Chumakov AM, et al. A novel, myeloid transcription factor, C/EBP epsilon, is up-regulated during granulocytic, but not monocytic, differentiation. *Blood*. 1997;90:2591-2600.
- Yamanaka R, Barlow C, Lekstrom-Himes J, et al. Impaired granulopoiesis, myelodysplasia, and early lethality in CCAAT/enhancer binding protein epsilon-deficient mice. *Proc Natl Acad Sci U S A*. 1997;94:13187-13192.
- Nielsen AL, Ortiz JA, You J, et al. Interaction with members of the heterochromatin protein 1 (HP1) family and histone deacetylation are differentially involved in transcriptional silencing by members of the TIF1 family. *EMBO J*. 1999;18:6385-6395.
- Anderson KL, Smith KA, Perkin H, et al. PU.1 and the granulocyte- and macrophage colony-stimulating factor receptors play distinct roles in late-stage myeloid cell differentiation. *Blood*. 1999;94:2310-2318.
- Gong A, Obermaler C, Boguth G, Weiss W. Recent developments in two-dimensional gel electrophoresis with immobilized pH gradients: wide pH gradients up to pH 12, longer separation distances and simplified procedures. *Electrophoresis*. 1999;20:712-717.
- Wilusz J, Pettine SM, Shenk T. Functional analysis of point mutations in the AAUAAA motif of the SV40 late polyadenylation signal. *Nucleic Acids Res*. 1989;17:3899-3908.
- Jansen R, Gerstein M. Analysis of the yeast transcriptome with structural and functional categories: characterizing highly expressed proteins. *Nucleic Acids Res*. 2000;28:1481-1488.
- Tamayo P, Satoru D, Mesirov J, et al. Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc Natl Acad Sci U S A*. 1999; 96:2907-2912.
- Proteometrics. ProFound search engine Web site. Available at: [http://www.proteometrics.com/profound\\_bin/WebProFound.exe](http://www.proteometrics.com/profound_bin/WebProFound.exe). Accessed May 4, 2001.

## Appendix

Gene symbols used in tables: Actb: actin, beta, cytoplasmic; Actg: actin, gamma, cytoplasmic; Actx: melanoma X-actin; Aldol: aldolase 1, A isoform; Arf5: ADP-ribosylation factor 5; Atf1: activating transcription factor 1; Atf2: activating transcription factor 2; Bcl2: basic transcription factor 3a; Bcrp: peripheral-type benzodiazepine receptor; C5: complement component 5, receptor 1/G protein-coupled receptor (C5a); Ccna2: cyclin B2; Cd36: CD36 antigen (collagen type I receptor, thrombospondin receptor)-like 2; Cd53: CD53 antigen; Cebpa: CCAAT/enhancer binding protein C/EBP, alpha; Cebpb: CCAAT/enhancer binding protein (C/EBP), beta; Cebpdl: CCAAT/enhancer binding protein (C/EBP), delta; Cebppe: CCAAT/enhancer binding protein (C/EBP), epsilon; Cfil: cofilin 1, nonmuscle; Cmkar4: chemokine (C-X-C) receptor 4; Cmkbr1: chemokine (C-C) receptor 1/Mip1a receptor; Cnlf: ciliary neurotrophic factor/zinc finger protein PZF; Copa: coatomer protein complex subunit alpha; Cpa3: carboxypeptidase A3, mast cell; Cr2: complement receptor 2; Ctrr: corticotropin releasing hormone receptor; Crry: complement receptor-related protein; Csf1r: CSF 1 (M-CSF) receptor/c-fms/CD115; Csf2ra: CSF 2 (GM-CSF) receptor, alpha, low-affinity/CD116; Csf2rb1: CSF 2 (GM-CSF) receptor, beta 2, low-affinity/IL-3 receptor-like protein (AIC2B)/CDw131;

Csf2rb2: CSF 2 (GM-CSF) receptor, beta 2, low-affinity/IL-3 receptor (AIC2A); Ctsb: cathepsin B; Ctsc: cathepsin C; Ctse: cathepsin E; Csgg: cathepsin G; Csh1: cathepsin H; Csl: cathepsin L; Cst: cathepsin S; Cybb: cytochrome b-245, beta; Ddr2: dopamine receptor 2; E2f1: E2F transcription factor 1; Ear2: eosinophil-associated ribonuclease 2; Ebi3: Epstein-Barr virus-induced gene 3/cytokine receptor-like molecule (EBI3); Ebf2: Balb/c neutrophil elastase; Ela2: elastase 2; Enh: enhancer of rudimentary homolog (Drosophila); Etv6: ethanol induced 6/steroid regulatory element binding transcription factor 1 (SREBF1) homolog; F2rl2: coagulation factor II (thrombin) receptor-like 2; Fcgr1g: Fc receptor, IgE, high affinity I, gamma polypeptide; Fcgr2b: Fc receptor, IgG, low affinity IIb; Fcgr3: Fc receptor, IgG, low affinity III; Fpr1: formyl peptide receptor 1/MLP receptor; Gabpb1: GA repeat binding protein (GABP-beta1 subunit); Gata2: GATA-binding protein 2; Gnas: guanine nucleotide binding protein, alpha stimulating; Gnb2-rs1: guanine nucleotide binding protein, beta-2, related sequence 1; Gpx3: glutathione peroxidase 3; Grg: related to Drosophila groucho gene; Grid1: glutamate receptor channel subunit delta 1; Grn: granulin; Gstm1: glutathione-S-transferase, mu 1; Gus-s: beta-glucuronidase structural; Gys3: glycogen synthase 3, brain; H2-D: histocompatibility 2, D

region locus 1; Hist2: histone gene complex 2; Hist5-2ax: H2A histone family, member X; Hmg1: high mobility group protein 1; Hsp60: heat shock protein, 60 kDa; Htr5a: 5-hydroxytryptamine (serotonin) receptor 5A; Id1: inhibitor of DNA binding 1/helix-loop-helix DNA binding protein regulator (Id); Id2: inhibitor of DNA binding 2; Ifng: interferon gamma receptor; Ifng2: interferon gamma receptor 2; Ii: Ia-associated invariant chain; Il1a: IL1 alpha; Il1r2: IL1 receptor, type II; Il2rg: IL2 receptor, gamma chain; Il4ra: IL4 receptor, alpha; Il10rb: IL10 receptor, beta; Il17r: IL17 receptor; Irf1: interferon regulatory factor 1; Irf2: interferon regulatory factor-2; Itgb2: integrin beta 2 (Cd18); Itp5: inositol 1,4,5-trisphosphate receptor (type 2); Jund1: Jun proto-oncogene-related gene 1/transcription factor JUN-D; Klf2: Kruppel-like factor LKLF; L-CCR: lipopolysaccharide inducible C-C chemokine receptor-related; Lcn2: lipocalin 2; Ldlr: low density lipoprotein receptor; Lsp1: Lymphocyte-specific 1/S37/pp52; Lst1: leucocyte-specific transcript 1; Ltb4r: leukotriene B4 receptor; Ltr: lymphotxin-beta receptor; Ltf: lactotransferrin; Ly6a: lymphocyte antigen 64; Ly6e: lymphocyte antigen 6 complex, locus E; Lyl1: lymphoblastic leukemia/bHLH factor; Lyzs: lysozyme; M6pr: mannose-6-phosphate receptor, cation dependent; Mad: Max dimerization protein; Man2c1: mannosidase, alpha, class 2C, member 1; Max: Max protein; Maz: MYC-associated zinc finger protein (purine-binding transcription factor); MBP: eosinophil granule major basic protein precursor; Mcpt8: mast cell protease 8; Mll: myeloid/lymphoid or mixed-lineage leukemia; Mmp13: matrix metalloproteinase 13/collagenase; Mmp9: matrix metalloproteinase 9/gelatinase B; Mpo: myeloperoxidase; Myb: myeloblastosis oncogene; Mybl2: myeloblastosis oncogene-like 2; Myc: myelocytomatosis oncogene; Myln: myosin light chain, alkali, nonmuscle; Nfatc2: nuclear factor of activated T cells, cytoplasmic 2; Nfe2: nuclear factor, erythroid-derived 2, 45 kDa; Nfkb1: NF-kappa-B (p105); Ngp: neutrophilic granule protein; NMDRGB: N-methyl-D-aspartate receptor glutamate-binding chain homolog; Npm1: nucleophosmin 1; Nr4a1: nuclear receptor subfamily 4, group A, member 1; Osi: oxidative stress induced; P2rx1: purinergic receptor P2X, ligand-gated ion channel, 1; P2ry2: purinergic receptor P2Y, G-protein-coupled 2; P40-8: P40-8, functional/laminin receptor; Pbx1: pre B-cell leukemia transcription factor 1; Pic: properdin factor, complement; Pira1: paired-Ig-like receptor A1; Pira5: paired-Ig-like receptor

A5; Pira6: paired-Ig-like receptor A6; Pirb: paired-Ig-like receptor B; Plaur: urokinase plasminogen activator receptor; PML: putative receptor protein (SP: P17152); Pml: promyelocytic leukemia; Prg: proteoglycan, secretory granule; Prg3: proteoglycan 3/eosinophil major basic protein 2; Prtn3: proteinase 3; Psma2: proteasome (prosome, macropain) subunit, alpha type 2; Ptmb4: prothymosin beta 4; Ptpre: protein tyrosine phosphatase, receptor type, C; Rac2: RAS-related C3 botulinum substrate 2; Rang: retinoic acid receptor, gamma; Rela: avian reticuloendotheliosis viral (v-rel) oncogene homolog A/NF-kappa-B p65; Rpl19: ribosomal protein L19; RPL8: ribosomal protein L8; Rps6a1: ribosomal protein S6 kinase polypeptide 1; Rps8: ribosomal protein S8; Rtu3: reticulon 3; S100a8: S100 calcium binding protein A8 (calgranulin A); S100a9: S100 calcium-binding protein A9 (calgranulin B); Sdf2: stromal cell-derived factor receptor 2; Sell: selectin L (lymphocyte adhesion molecule 1); Sema4d: semaphorin 4D; Sepp1: selenoprotein P, plasma, 1; Sfp1: SFFV proviral integration 1; Shfdg1: split hand/foot deleted gene 1; Slc10a1: solute carrier family 10 (sodium/bile acid cotransporter family), member 1; Slpi: secretory leukocyte protease inhibitor; Sox15: SRY-box containing gene 15; Spi2-1: serine protease inhibitor 2-1; Srf1: scavenger receptor class B1; Stat3: signal transducer and activator of transcription 3; Stat5a: signal transducer and activator of transcription 5A; Stat6: signal transducer and activator of transcription 6; Stra14: basic-helix-loop-helix protein-retinoic acid induced; Tbx1: TBX1 protein/LPS-induced TNF-alpha factor homolog; Tcrb: T-cell-receptor germline beta-chain gene constant region; Tcrb-V4: T-cell-receptor gamma, variable 4; Tetex1: t-complex testis expressed 1; Tfdp1: transcription factor Dp 1; Tif1b: transcriptional intermediary factor 1, beta; Tlr4: toll-like receptor 4; Tnfisf1a: TNF receptor superfamily, member 1a; Tnfisf1b: TNF superfamily, member 1b; Tomm70a: translocase of outer mitochondrial membrane 70 (yeast) homolog A; Tpi: triosephosphate isomerase; Trp53: transformation-related protein 53; Ubb: ubiquitin B; Usf2: upstream transcription factor 2; Ybx1: Y box transcription factor; Ybx3: Y box binding protein; Zfp11-6: zinc finger protein 11-6; Zfp18: zinc finger protein 18 homolog; Zfp36: zinc finger protein 36; Zfp162: zinc finger protein 162; Zfp216: zinc finger protein 216; Zfpml: zinc finger protein, multitype 1; Znfn1a1: zinc finger protein, subfamily 1A, 1 (Ikaro); Zyx: zyxin.

## A Genomic and Proteomic Analysis of Activation of the Human Neutrophil by Lipopolysaccharide and Its Mediation by p38 Mitogen-activated Protein Kinase\*

Received for publication, January 24, 2002  
Published, JBC Papers in Press, April 9, 2002, DOI 10.1074/jbc.M200755200

Michael B. Fessler<sup>‡§</sup>, Kenneth C. Malcolm<sup>§</sup>, Mark William Duncan<sup>¶</sup>, and G. Scott Worthen<sup>‡§¶</sup>

From the <sup>‡</sup>Department of Medicine, Division of Pulmonary Sciences and Critical Care Medicine, University of Colorado Health Sciences Center, the <sup>§</sup>Department of Medicine, National Jewish Medical and Research Center, and the <sup>¶</sup>Biochemical Mass Spectrometry Facility, School of Pharmacy, University of Colorado Health Sciences Center, Denver, Colorado 80262

Bacterial lipopolysaccharide (LPS) evokes several functional responses in the neutrophil that contribute to innate immunity. Although certain responses, such as adhesion and synthesis of tumor necrosis factor- $\alpha$ , are inhibited by pretreatment with an inhibitor of p38 mitogen-activated protein kinase; others, such as actin assembly, are unaffected. The aim of the present study was to investigate the changes in neutrophil gene transcription and protein expression following lipopolysaccharide exposure and to establish their dependence on p38 signaling. Microarray analysis indicated expression of 13% of the 7070 Affymetrix gene set in nonstimulated neutrophils, and LPS up-regulation of 100 distinct genes, including cytokines and chemokines, signaling molecules, and regulators of transcription. Proteomic analysis yielded a separate list of up-regulated modulators of inflammation, signaling molecules, and cytoskeletal proteins. Poor concordance between mRNA transcript and protein expression changes was noted. Pretreatment with the p38 inhibitor SB203580 attenuated 23% of LPS-regulated genes and 18% of LPS-regulated proteins by  $\geq 40\%$ . This study indicates that p38 plays a selective role in regulation of neutrophil transcripts and proteins following lipopolysaccharide exposure, clarifies that several of the effects of lipopolysaccharide are post-transcriptional and post-translational, and identifies several proteins not previously reported to be involved in the innate immune response.

Lipopolysaccharide (LPS),<sup>1</sup> a component of the outer cell wall of Gram-negative bacteria, evokes a variety of functional responses in the human neutrophil (PMN) after binding to a plasma membrane receptor complex that involves the Toll-like

receptors (TLRs) (1–5). These “immediate” functional responses, including actin assembly, adhesion, activation of nuclear factor-kappa B (NF- $\kappa$ B), and priming for an enhanced secretory response and for release of reactive oxygen intermediates, appear to be central both to the innate immune response and to the pathogenesis of several inflammatory human diseases, including sepsis and the acute respiratory distress syndrome (6). p38 mitogen-activated protein kinase (p38 MAPK) has been shown to mediate LPS-induced PMN adhesion, NF- $\kappa$ B activation, and TNF- $\alpha$  and IL-8 translation and release (7), and its blockade attenuates LPS-induced PMN accumulation in the airspace (8). However, other cascades almost certainly lead to downstream effectors of the LPS signal; for example, actin assembly appears to be p38 MAPK-independent (9). An improved understanding of the transcriptional and translational responses of the neutrophil to LPS and the modulation of these responses by p38 MAPK might carry pathogenetic and therapeutic implications.

Historically, it has been believed that the downstream PMN transcriptional response to LPS is static and that PMN functional responses to LPS that depend on *de novo* protein synthesis are primarily limited to the release of cytokines (10). However, recent studies indicate a robust transcriptional response (11). To date, most studies have relied upon and reported a short list of functional assays of the LPS-exposed PMN; therefore, no exhaustive investigation of either the transcriptional response or protein synthetic repertoire of the PMN has been reported. Although several techniques have been used to evaluate transcripts, the screening of global changes in mRNA by microarray analysis has only recently become possible. In this way, thousands of genes can be screened in an unbiased fashion for transcript abundance. Such genomic screens in mammalian cells have previously been applied to define altered expression profiles in response to agonists (12) and to drug action (13) and during cell cycle progression (14).

Although DNA microarray technology is expected to provide insight into the response of the human PMN to LPS (15), inhibition of LPS-stimulated IL-1 and TNF- $\alpha$  production by p38 MAPK inhibitors in THP-1 cells (16) and of TNF- $\alpha$  synthesis in human PMNs (9) occurs at a translational level and would therefore not be detected by DNA microarrays. Furthermore, in other systems, such as yeast and human liver, mRNA and protein levels show poor correlation (17, 18). Proteomics is a complementary tool for assessing global changes in cellular protein expression, thereby providing additional insight into cellular signal regulation. A proteomic approach has proven useful in different systems for dissecting signal transduction cascades and describing their output (19, 20) and has even

\* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

<sup>‡</sup> To whom correspondence should be addressed: Dept. of Medicine, D403, Neustadt Bldg., National Jewish Medical and Research Center, 1400 Jackson St., Denver, CO 80206. Tel.: 303-398-1171; Fax: 303-398-1381; E-mail: worthens@njc.org.

<sup>1</sup> The abbreviations used are: LPS, lipopolysaccharide; DTT, dithiothreitol; IEF, isoelectric focusing; IFN, interferon; IL, interleukin; MALDI-TOF, matrix-assisted laser desorption/ionization-time of flight; MAPK, mitogen-activated protein kinase; NF- $\kappa$ B, nuclear factor-kappa B; pI, isoelectric point; PMN, neutrophil (polymorphonuclear leukocyte); TLR, Toll-like Receptor; TNF, tumor necrosis factor; CHCA,  $\alpha$ -cyano-4-hydroxycinnamic acid; AEBSEF, 4-(2-aminoethyl)-benzenesulfonylfluoride hydrochloride; MS, mass spectrometry; CaM, Ca<sup>2+</sup>/calmodulin; ERK, extracellular signal-regulated kinase; E-64, epoxysuccinyl-64.

recently been used to detect novel upstream messengers involved in LPS signal transduction (21). We have applied DNA microarrays and proteomics to define and compare transcriptional and post-transcriptional alterations in the LPS-exposed PMN and to establish the dependence of these alterations on p38 MAPK signaling.

#### EXPERIMENTAL PROCEDURES

**Materials**—Endotoxin-free reagents and plastics were used in all experiments. Aprotinin, leupeptin, AEBSF, E-64, pepstatin, and bestatin protease inhibitors, spermine HCl, and  $\alpha$ -cyano-4-hydroxycinnamic acid (CHCA) were all purchased from Sigma Chemical Co. (St. Louis, MO). SB203580, a p38 MAPK inhibitor, was purchased from Calbiochem-Novabiochem Corp. (San Diego, CA). For two-dimensional PAGE, rehydration buffer, equilibration buffers, vertical electrophoresis solutions, and 10% homogeneous polyacrylamide slab gels were purchased from Genomic Solutions, Inc. (GSI, Ann Arbor, MI). Sequencing grade porcine trypsin was purchased from Promega (Madison, WI).

**LPS Incubation**—PMNs were isolated by the plasma Percoll method (22), a technique that yields less than 5% monocytic contamination, and resuspended at a concentration of  $15.4 \times 10^6$ /ml in RPMI 1640 culture medium (BioWhittaker, Walkersville, MD) supplemented with 10 mM HEPES (pH 7.6) and 1% heat-inactivated platelet-poor plasma. After addition of 100 ng/ml *Escherichia coli* 0111:B4 LPS (List Biological), incubation was carried out with continuous rotation (4 h, 37 °C) both in the presence and absence of SB203580. Both Affymetrix analysis and proteomic analysis utilized  $75 \times 10^6$  cells. For microarray analysis, nonstimulated and 4-h-treated PMNs were collected from three separate donors. A more detailed time course following LPS exposure was performed using polymerase chain reaction. For proteomic analysis, LPS incubations from separate donors ( $n = 6$ ) were performed and then analyzed individually. Control and post-LPS incubation PMNs were washed (0.34 M sucrose/1 mM EDTA/10 mM Tris) and then lysed in a modified rehydration buffer (GSI, Ann Arbor, MI) supplemented with 2 M thiourea, 50 mM dithiothreitol (DTT), 22.5 mM spermine HCl, and a mixture of six protease inhibitors (10  $\mu$ g/ml aprotinin, 10  $\mu$ g/ml leupeptin, 2 mM AEBSF, 5  $\mu$ g E-64, 1  $\mu$ g pepstatin, 10  $\mu$ g bestatin). DNA was pelleted by centrifugation at  $250,000 \times g$  for 60 min (23).

**Affymetrix Oligonucleotide Array**—Five micrograms of total RNA was isolated with TRIzol (Invitrogen) and RNeasy columns (Qiagen) and subsequently labeled with biotin as described by Affymetrix. Briefly, first-strand synthesis was accomplished with Superscript II reverse transcriptase (Invitrogen) using a T7-oligo(dT)<sub>24</sub> primer for 1 h at 42 °C followed by second-strand synthesis using *E. coli* DNA polymerase I and RNase H (Invitrogen) at 16 °C for 2 h. Double-stranded DNA was used as a template for *in vitro* transcription with T7 RNA polymerase in the presence of biotin-labeled UTP and CTP using the BioArray High Yield RNA transcript labeling kit (Enzo). Fifteen micrograms of cRNA was fragmented and used for hybridization to Affymetrix HuGene 6800FL Genechips. Each sample was hybridized initially using a Test2 Genechip to test for sample degradation and full-length *in vitro* translation. Data were analyzed using Affymetrix Genechip software. Results from three separate donors were analyzed.

**Reverse Transcription and Polymerase Chain Reaction**—cDNA was prepared by reverse transcription using 2  $\mu$ g total RNA, derived from  $20 \times 10^6$  cells that were treated as indicated. Polymerase chain reactions were performed using specific primers for *Mx-1*, *TNF- $\alpha$* , *MCP-1*, *p65*, *S100A4*, and glyceraldehyde-3-phosphate dehydrogenase.

**Two-dimensional PAGE**—The protein concentration of the lysates was measured as described by Bradford *et al.* (24). Poor isoelectric focusing (IEF) results were encountered unless the polycationic spermine was diluted (data not shown); therefore, lysates were diluted with rehydration buffer (GSI, Ann Arbor, MI) to achieve a final spermine concentration of 6 mM. Equal protein loads (1.5 mg) of control and LPS-stimulated neutrophils were used to rehydrate IEF gels overnight (18 cm, pH 3–10 nonlinear Immobiline DryStrip IEF gels, Amersham Biosciences, Piscataway, NJ). IEF was performed at 20 °C to 100-kVh (Phaser, GSI) under mineral oil, followed by two 10-min SDS equilibration steps (DTT and then iodoacetamide-containing equilibration buffers, GSI) and then by vertical electrophoresis on 10% homogeneous polyacrylamide slab gels (GSI) at 500 V. Protein spots were visualized by agitation in colloidal Coomassie Brilliant Blue G-250 (16 h) (25), followed by destaining in deionized water (20 h). In separate experiments, control and LPS-stimulated PMN lysates from three donors were pooled and then analyzed by two-dimensional PAGE using overlapping narrow isoelectric point (pI) ranges (18 cm, pH 5.0–6.0, 5.6–

6.7, and 6–11, Amersham Biosciences, Piscataway, NJ). Identical IEF and vertical electrophoresis parameters were used for all gels.

**Image Analysis of Two-dimensional Gels**—Colloidal Coomassie-stained gels were digitized using a Powerlook II (UMAX Data Systems, Inc., Taiwan) flatbed scanner with 8-bit dynamic range and 150-dpi resolution. BioImage (GSI, Ann Arbor, MI) 2D-Analyzer software was used to locate, quantitate, and match protein spots on the control and LPS gel images. Analysis was performed by assigning 50 common anchor spots between paired images; the remaining spots were compared by a constellation-matching algorithm. All data were then carefully reviewed by the operator to account for any discrepancies. Protein loading between control and experimental gels may have varied because of inconsistencies in rehydration of the different IEF gel strips; therefore, gel images were normalized so that the sum of the integrated intensities of all matched spots on paired gels was made equal. Control and LPS-stimulated gel images from individual donor experiments were matched to generate composite images; composite images were then matched into a master composite image to track the LPS response of protein spots among different donors (26). Only those spots that were common (image-matched) to all original 12 (pH 3.0–10.0) gels were considered for further analysis. For these spots, the LPS-induced change in integrated intensity in the six experiments was subjected to statistical analysis with a two-tailed Student's *t* test, and those spots with  $p < 0.05$  were identified by peptide mass fingerprinting (described below). For the narrow range (pH 5.0–6.0, 5.6–6.7, and 6–11) two-dimensional PAGE experiments using pooled donors, only those spots with concordant regulation exceeding 1.5-fold or that appeared *de novo* in the LPS gel in two repeat experiments were further analyzed.

**In-gel Tryptic Digestion**—In-gel digestion of protein spots was performed with sequencing grade porcine-modified trypsin using the method of Hellman *et al.* (27). Tryptic peptides were then extracted (50  $\mu$ l of 50% acetonitrile/5% trifluoroacetic acid, 2 h), and the supernatant was taken to dryness in a vacuum centrifuge and then redissolved in trifluoroacetic acid (20  $\mu$ l, 0.5%). Peptides were then purified and concentrated using ZipTip<sub>C18</sub> pipette tips (Millipore, Bedford, MA).

**MALDI-TOF Mass Spectrometry**—Analyses were performed on an Applied Biosystems matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) Voyager-DE PRO mass spectrometer (Framingham, MA) operated in delayed extraction mode. Samples (0.5  $\mu$ l) were spotted onto a sample plate to which matrix (0.5  $\mu$ l of 10 mg/ml CHCA) was added. The sample-matrix mixture was dried at room temperature and then analyzed in reflector mode. CHCA was also spotted alone as a negative control. Spectra were the sum of 100 laser shots, and those peaks with a signal-to-noise ratio of greater than 3:1 were selected for data base searching. Spectra were internally calibrated using autolytic trypsin peptides ( $m/z$  842.51, 2211.10).

**Data Base Searching Algorithm**—The monoisotopic masses for each protonated peptide were: (a) entered into the program MS-Fit (available at [prospector.ucsf.edu](http://prospector.ucsf.edu)) for searches against the Swiss-Prot, NCBI, and GenPept databases, and (b) entered into Mascot (available at [matrixscience.com](http://matrixscience.com)), an algorithm testing statistical significance of peptide mass fingerprinting identifications. For MS-Fit searches, masses derived from trypsin, CHCA, keratin, and Coomassie Brilliant Blue G-250 were excluded. Search parameters included a maximum allowed peptide mass error of 0.1 Da (0.8 Da in the few instances in which linear mode was used), consideration of one incomplete cleavage per peptide, pI range of 3.0–10.0, and molecular mass range of 1–200 kDa. Accepted modifications included carbamidomethylation of cysteine residues (from iodoacetamide exposure following IEF) (28) and methionine oxidation, a common modification occurring during SDS-PAGE (29). Protein identifications were assigned when three criteria were met: 1) statistical significance ( $p < 0.05$ ) of the match when tested by Mascot ([matrixscience.com](http://matrixscience.com)); 2) >20% sequence coverage by the tryptic peptides; and 3) concordance ( $\pm 16\%$ ) with the molecular weight and pI of the parent two-dimensional PAGE protein spot. The following special exceptions were considered: (a) protein identifications not fulfilling criterion 2 were still assigned if criteria 1 and 3 were fulfilled and no other *Homo sapiens* proteins with peptide mass-matched  $p$  values < 0.05 were identified by Mascot; (b) if criterion 3 was not fulfilled (lower than expected molecular weight), a cleavage product of the identified protein was inferred, and the cumulative molecular weight of the tryptic peptides was compared with that of the two-dimensional-PAGE spot to ensure that it was not exceeded; (c) if criterion 3 was not fulfilled (isolated discordance between theoretical and observed pI), post-translational modification of an unrecovered peptide was inferred; and (d) if two or more *H. sapiens* protein assignments with >4 mutually exclusive matching peptides were identified, a protein mixture in the two-dimensional PAGE

spot was inferred and further analysis halted (quantitative conclusions regarding the individual protein constituents could not be drawn).

## RESULTS

**Genes Differentially Expressed in LPS-stimulated Neutrophils**—Human PMNs were left untreated or incubated in the presence of 100 ng/ml LPS for 4 h. As a control to confirm that the PMNs were quiescent at baseline and that LPS resulted in normal stimulation, mRNA was isolated, cDNA was prepared, and PCR for TNF- $\alpha$  was performed. Little TNF- $\alpha$  expression was seen in nonstimulated cells, whereas LPS treatment led to an increase in expression in each of the donors subsequently used for microarray analysis (data not shown). No macrophage-colony stimulating factor receptor transcript was detected by oligonucleotide microarray analysis, confirming there was no significant monocytic contamination.

Human PMNs express a limited repertoire of mRNA transcripts at baseline but respond to LPS with differential expression of genes in many families. Considering only those genes present by microarray analysis in all three donors, unstimulated PMNs expressed 13.0% (923 of 7070 genes) of the Affymetrix gene set. Gene classes represented at baseline include metabolic enzymes, structural proteins, receptors, signaling proteins, and transcription factors. By comparison, human monocytes expressed ~40% and human fibroblasts ~35% of the represented genes (data not shown). By the criterion of a >3-fold increase in expression in all three donors on Affymetrix oligonucleotide array analysis, exposure of PMNs to LPS for 4 h resulted in the up-regulation of 100 genes (Table I).

Genes from several different functional classes were induced in PMNs following LPS exposure. Of interest, a number of transcriptional regulators were induced, including transcription factors of the NF- $\kappa$ B family. The transcriptional NF- $\kappa$ B complex has previously been implicated in the regulation of the genes induced by LPS (11). The genes for several cytokines and chemokines were also found to be up-regulated. These include TNF- $\alpha$ , IL-1 $\beta$ , IL-6, MCP-1, MIP-3 $\alpha$ , and MIP-1 $\beta$  (Table I). PCR was performed to confirm the results from the microarray analysis. PCR analysis on selected genes indicates that the time course for changes can be rapid or delayed but parallel the changes found in the array at the 4-h time point (data not shown). Other up-regulated genes included those for metabolic enzymes, immune response molecules, kinases, phosphatases, signaling molecules, adhesion and cytoskeletal components, interferon-stimulated genes, and those with unknown or miscellaneous function (Table I).

LPS stimulation of PMN also resulted in the down-regulation of 56 genes (Table II). Down-regulated genes were identified as transcriptional regulators, protein and lipid kinases and phosphatases, structural molecules, and signaling molecules. Genes for metabolic proteins were also evident, as were several uncharacterized genes.

**Two-dimensional PAGE and Image Analysis**—In contrast to the limited number of transcripts found at baseline, PMNs were found to express a large number and variety of proteins in the nonstimulated state (Fig. 1, A and C, and Tables III–V). Reproducible protein expression patterns were found on the pH 3.0–10.0 gels, and the majority of proteins fell in the pH 5.0–7.0 range (Fig. 1A). The basic region (pH > 7.0) consistently exhibited poor resolution, precluding meaningful image analysis and further workup (data not shown). Depending on the spot-finding parameters (minimum spot intensity, filter width) selected on the image analysis software, spot-by-spot manual editing was found to be necessary to avoid over- and under-detected spots; moreover, further manual editing was performed to screen for unmatched and mismatched spots following matching of paired control and LPS-stimulated gels. After spot

editing, ~1200 well-resolved spots were evident on each pH 3.0–10.0 gel. In an attempt to improve resolution of the pH range bearing the greatest number of well-resolved spots, overlapping narrow pH range gels (pH 5.0–6.0, 5.5–6.7, 6–11) were also run. Of interest, a similar number of well-resolved spots (~1200) were detected on the narrow pH range gels (Fig. 1, C and D). Assuming a detection limit for Coomassie of 15 ng (0.25 pmol, or  $1.5 \times 10^{11}$  molecules, for a 60-kDa protein) and a protein load per gel corresponding to  $75 \times 10^6$  PMNs, we estimate a detection limit on our gels of 2000 molecules/cell for a 60-kDa protein. As investigators have suggested in other cell lines with the use of high resolution two-dimensional-PAGE methods (30), we estimate that >10,000 proteins are expressed in the resting PMN.

Human PMNs respond to LPS with the differential expression of a large number of proteins. In the six individual pH 3.0–10.0 experiments, the number of protein spots that increased in integrated intensity by at least 50% following LPS exposure was 185, 122, 104, 104, 96, and 131, respectively. The number of protein spots that decreased by at least 50% following LPS exposure was 72, 151, 102, 98, 128, and 97, respectively. Although gel-to-gel regional variability in resolution was expected to account for individual spots not being well visualized on particular gels, only those spots that were matched to all 12 original gels were analyzed further. Overall, the number of spots matched to all 12 original gels was 125. The numbers of spots that were both matched to all 12 original gels and that increased by at least 50% in integrated intensity in the individual experiments following LPS exposure were 46, 13, 17, 27, 22, and 20, respectively. The numbers of spots that were matched to all 12 gels and that decreased by at least 50% were 6, 22, 17, 22, 34, and 28, respectively. The LPS-induced change in integrated intensity of the 125 spots that were matched to all 12 original gels was subjected to statistical analysis with a two-tailed Student's *t* test, and those spots with statistically significant ( $p < 0.05$ ) regulation among the six experiments were identified by peptide mass fingerprinting (Table III).

**Identification of LPS-regulated Proteins**—Several proteins were consistently up-regulated on the pH 3.0–10.0 gels (Table III), including regulators of inflammation (annexin III) and signaling molecules (Rab-GDP dissociation inhibitor  $\beta$ ). Several actin fragments were seen to be consistently up-regulated in the six experiments following LPS exposure (Table III). Of interest, the proteasome  $\beta$  chain was also consistently up-regulated. Down-regulated proteins included other signaling molecules, such as Rho GTPase activating protein 1.

On the pH 5.0–6.0 and 5.5–6.7 gels, several proteins were found to show increases of greater than 1.5-fold following LPS exposure (Tables IV and V), including cytoskeletal proteins, such as moesin, nonmuscle myosin heavy chain, and a putative phosphorylated form of nonmuscle myosin heavy chain, and signaling molecules, such as protein phosphatase 1 and PO $_4$ -stathmin. The putative phosphorylated form of nonmuscle myosin heavy chain (spot #1101) was positioned 0.03 pH unit more acidic than the unmodified protein (spot #1102) (Fig. 1D) and was distinguished by a tryptic peptide (*m/z* 1366.74) not present in the unmodified protein, consistent with phosphorylation of serine 685. Serine 685 is predicted by NetPhos 2.0 Prediction Server (available at [www.cbs.dtu.dk/services/NetPhos/\(31\)](http://www.cbs.dtu.dk/services/NetPhos/(31))) to be a high probability phosphorylation residue and by ScanProsite ([www.expasy.ch/tools/scnpsite.html](http://www.expasy.ch/tools/scnpsite.html)) to be a substrate for protein kinase C. The tryptic phosphopeptide identified in PO $_4$ -stathmin, extending from residues 15 to 27 (1468.7 Da), is consistent with phosphorylation of either serine 16, a known substrate for Ca $^{2+}$ /calmodulin (CaM)-dependent kinases (32), or serine 25, a known substrate for p38 $\delta$  and ERK (Fig. 2A)

TABLE I  
Human neutrophil genes induced after 4 h of LPS exposure

Description	GenBank™ no.	Change-fold
<b>Transcriptional regulation</b>		
<i>Pleiomorphic adenoma gene-like 2</i>	D83784	16.8
<i>NFKB2</i>	S76638	12.3
<i>NFKB1E</i>	U91616	11.5
<i>p65</i>	L19067	8.4
<i>BCL3</i>	U05681	7.7
<i>X-box binding protein 1</i>	M31627	7.5
<i>Metal-regulatory transcription factor 1</i>	X78710	7.4
<i>Ets-2</i>	J04102	7.4
<i>c-Rel</i>	X75042	6.2
<i>NFKB1</i>	M58603	5.8
<i>Basic leucine zipper transcription factor, ATF-like</i>	U15460	4.7
<i>IKB</i>	M69043	3.8
<i>MAX dimerization protein</i>	L06895	3.6
<i>DIF2</i>	S81914	3.1
<b>Cytokines and receptors</b>		
<i>MCP-1</i>	M69203	78.7
<i>MIP-1<math>\beta</math></i>	M72885	48.8
<i><math>\alpha</math>Helix coiled-coil rod homolog</i>	AF014958	20.8
<i>IL-1<math>\beta</math></i>	X04500	17.6
<i>GRO3 (beta)</i>	M57731	17.3
<i>TNF-<math>\alpha</math></i>	X02910	14.5
<i>MIP-3<math>\alpha</math></i>	U64197	8.1
<i>IL10RA</i>	U00672	7.3
<i>IL-6</i>	Y00081	6.3
<i>GR<math>\alpha</math></i>	X54489	4
<i>HM74</i>	D10923	3.8
<b>Immune response</b>		
<i>Orosomucoid</i>	X02544	20.2
<i>Complement component C3</i>	K02765	12.8
<i>Protease inhibitor 9</i>	U71364	9.5
<i>Complement component 3a receptor 1</i>	U28488	6.1
<i>Protease inhibitor 3</i>	L10343	4.9
<i>SLPI/antileukoprotease</i>	X04470	4.7
<i>ELANH2/elastase inhibitor</i>	M93056	4.6
<i>CD58</i>	Y00636	3.8
<i>Complement component PFC</i>	M83652	3.5
<b>Kinases</b>		
<i>CNK/FNK/PLK-like</i>	U56998	16.2
<i>Cot</i>	D14497	11.9
<i>Pim-2</i>	U77735	9.5
<i>LMK2</i>	D45906	4.3
<b>Phosphatases</b>		
<i>PAC-1/DUSP2</i>	L11329	11.8
<i>DUSP5</i>	U15932	5.3
<i>PHA1</i>	U73477	3.4
<b>Signaling molecules</b>		
<i>TNFAIP1/A20</i>	M59465	10
<i>TRAF1</i>	U19261	6.2
<i>RanBP2</i>	D42063	5.6
<i>GNA15</i>	M63904	5.2
<i>PTAFR</i>	D10202	3.9
<b>Adhesion and cytoskeleton</b>		
<i>ICAM1</i>	M24283	22.4
<i>CEACAM1 (biliary glycoprotein)</i>	X16354	6.3
<i>LIMS1</i>	U09284	6.1
<i>SNL/actin bundling protein</i>	U03057	5.9
<i>Galectin-1/LGALS1</i>	M57710	4.7
<i>MEMD/ALCAM</i>	U30999	4.2
<i>CD44</i>	HG2981—HT3125	3.9
<i>TSG-6</i>	M31165	3.7
<b>Metabolic</b>		
<i>GTP cyclohydrolase I</i>	U19523	13.5
<i>NDUFV2/ubiquinone reductase</i>	M22538	8.6
<i>PSMA6/proteasome iota</i>	X59417	8.4
<i>UDP-galactose transporter (SLC35A2)</i>	D84454	7.3
<i>PLAU (urokinase)</i>	X02419	6.4
<i>KYNUH/kynurenine hydrolase</i>	U57721	5.5
<i>AMPD3</i>	D12775	5
<i>P4HA1/prolyl 4-hydroxylase</i>	M24486	4.7
<i><math>\gamma</math> Glutamylcysteine synthetase</i>	L35546	4.5
<i>ATP6D</i>	J05682	4.2
<i>ATP6S1</i>	D16469	4

TABLE I—continued

Description	GenBank™ no.	Change-fold
<i>Glycerol kinase</i>	X68285	3.6
<i>PAC1</i>	L09229	3.5
<i>AK3</i>	X60673	3.3
Interferon-inducible		
<i>ISG15</i>	M13755	22.5
<i>Mx1</i>	M33882	19.4
<i>IFI56</i>	M24594	12.1
<i>INDO</i>	M34455	5.2
<i>GBPI</i>	M55542	4.3
<i>PRKR</i>	U50648	3.7
<i>IFIT4</i>	U52513	3.6
<i>IFI54</i>	M14660	3.5
<i>IFI58</i>	U34605	3.5
<i>IFP35</i>	U72882	3
Other		
<i>Gos2</i>	M72885	48.8
<i>MIHC/IIAP1</i>	U37546	7.2
<i>KIAA0105</i>	D14661	5.1
<i>KIAA0118</i>	D42087	5
<i>SNAP23</i>	U55936	5
<i>CASP5</i>	U28015	4.8
<i>KIAA0113</i>	D30755	4.9
<i>KIAA0255</i>	D87444	4.7
Hepatoma-derived GF	D16431	4.7
<i>PTGS2</i>	D28235	4.6
<i>CD48</i>	M37766	4.3
<i>UNC119 homolog</i>	U40998	4.2
<i>KIAA0151</i>	D63485	3.9
<i>Rab1b</i>	XM035660	3.8
<i>Annexin VII</i>	J04543	3.7
<i>KIAA0110</i>	D14811	3.7
<i>Adrenomedullin</i>	D14874	3.7
<i>AIM1</i>	U83115	3.6
<i>KIAA0250</i>	D87437	3.2
<i>PS-1</i>	L06175	3.2
Scavenger receptor expressed by endothelial cells	D63483	3.2
<i>VHL</i>	L15409	3.1

(33). Assuming that no other multiply phosphorylated stathmin species had escaped detection, analysis of the integrated intensities of the  $PO_4$ -stathmin and stathmin spots indicates that the percentage of the  $PO_4$  form of total cellular stathmin increased from 11% to 38% with LPS stimulation (Fig. 2B). This is similar to a previous report of an increase from <10% to 35–40% of the Ser<sup>25</sup>-phosphorylated form in Jurkat cells stimulated with anti-CD3 (34).

**Effect of SB203580 on LPS-stimulated Gene Expression—**Gene expression analysis of PMNs stimulated with LPS indicated that the majority of genes induced by LPS were unaffected by prior treatment of PMN with SB203580. Of the 100 genes up-regulated by LPS, the up-regulation of 23 was inhibited by greater than 40% (Table VI). The majority of these genes affected by SB203580 were inhibited by less than 60%, whereas only six were inhibited by greater than 80%, all of which represent previously identified interferon-stimulated genes. Induction of cytokine genes by LPS, with the exception of *IL-6*, was generally unaffected by SB203580.

**Effect of SB203580 on LPS-stimulated Protein Expression—**Similar to the effect of SB203580 on LPS-stimulated gene expression, little effect of SB203580 was seen on expression levels for the majority of LPS-regulated proteins (Table VII). Two exceptions are annexin III and  $\alpha$ -enolase, for which LPS-stimulated expression was attenuated in the presence of the p38 MAPK inhibitor.

**Comparison of Microarray and Proteomics Results—**Of the LPS-regulated proteins identified by peptide mass fingerprinting for which probes were present on the oligonucleotide microarray, poor concordance was found at the mRNA level (Table VIII). For 13 LPS-up-regulated proteins, 2 corresponding

mRNA transcripts were up-regulated, 1 was down-regulated, 5 were unchanged, and 5 were not detected by the Affymetrix chip. For 5 down-regulated proteins, 3 corresponding transcripts were down-regulated, 1 was unchanged, and 1 was not detected. Varying patterns of LPS regulation emerge for those candidates detected at both the transcript and protein level. Proteasome  $\beta$  chain was up-regulated at both the transcript and protein levels (Table VIII), with no notable effect of SB203580 on expression at either level. Similarly, CAP1, RhogAP1, and ficolin 1 were down-regulated at both the mRNA transcript and protein level (Table VIII), with no notable effect of SB203580. Annexin III was down-regulated at the transcript level and up-regulated at the protein level, with an inhibitory effect of SB203580 seen only at the protein level (Tables VII and VIII).

#### DISCUSSION

Interaction of bacterial LPS with the human PMN represents a model system for studying the activation and output of the innate immune system during infection and inflammation. A recent publication (35) describes the gene expression changes of a cultured monocytic cell line after infection by the Gram-positive bacterium *Listeria monocytogenes*. The cell wall components of Gram-positive bacteria, like Gram-negative-derived LPS (*i.e.* from *E. coli*), are known to signal through TLRs (36, 37). Importantly, many of the expression changes found in LPS-stimulated PMNs in the present study were also described in the bacteria-exposed monocytic cells, indicating that many of the gene expression changes seen in bacterial infection are likely mediated by TLRs (38, 39) and that the LPS model system accurately reflects exposure of immune cells to infec-

TABLE II  
Human neutrophil genes repressed (>4-fold) after 4 h of LPS exposure

Description	GenBank™ no.	Change
<b>Kinases</b>		<i>-fold</i>
<i>CAMK II, gamma</i>	U50360	-4
<i>Diacylglycerol kinase, delta</i>	D63479	-4.2
<i>PRKCL2/PRK2 protein kinase C-like 2</i>	U33052	-4.3
<i>MAPKAPK3</i>	U09578	-6.3
<i>Protein kinase Ht31, cAMP-dependent</i>	HG2167-HT2237	-8
<i>CAMK II</i>	L07044	-9.8
<b>Transporters</b>		
<i>SLC25A5/solute carrier family 25, member 5</i>	J02683	-4.2
<i>SLC19A1; folate transporter</i>	U17566	-4.4
<i>SLC2A3; facilitated glucose transporter</i>	M20681	-5
<b>Metabolic</b>		
<i>Carbonic anhydrase IV</i>	L10955	-4.4
<i>RNase A family, k6</i>	U64998	-4.5
<i>Glycogen phosphorylase; liver</i>	M14636	-4.6
<i>Inositol polyphosphate-5-phosphatase</i>	U67650	-4.6
<i>Inositol 1,3,4-trisphosphate 5/6-kinase</i>	U51336	-4.7
<i>Transketolase</i>	L12711	-4.8
<i>Protein phosphatase 4, reg. subunit 1 (clone 23840)</i>	U79267	-4.9
<i>Cytidine deaminase</i>	L27943	-5.4
<i>MGAT1</i>	M55621	-5.4
<i>HMOX1</i>	X06985	-5.4
<i>MAN2A2</i>	L28821	-5.8
<i>Glycogenin (also represents U31525)</i>	HG4334-HT4604	-5.9
<b>Structural</b>		
<i>Fibrinogen-like protein (pT49 protein)</i>	Z36531	-4.2
<i>H2AFZ</i>	M37583	-4.7
<i>Paxillin</i>	U14588	-4.9
<i>Lamin B R</i>	L25931	-5.9
<i>Dynamin 2</i>	L36983	-6.2
<i>Actinin 1</i>	M95178	-6.7
<i><math>\alpha</math>-Tubulin</i>	X01703	-10
<i>Tubulin, <math>\alpha</math>1, isoform 44</i>	HG2259-HT2348	-15
<b>Transcriptional regulators</b>		
<i>Lymphoblastic leukemia-derived sequence 1</i>	M22638	-4.4
<i>MAX-interacting protein 1</i>	L07648	-4.5
<i>Nuclear factor erythroid 2 isoform f</i>	S77763	-6
<i>Transducer of ERBB2, 1</i>	D38305	-6.9
<i>NFATC4</i>	L41067	-7.8
<i>ATF-2 (CRE-Bpa)</i>	L05515	-9.6
<b>Receptors</b>		
<i>Lymphotoxin <math>\beta</math> receptor</i>	L04270	-4.4
<i>Folate receptor 3 (gamma)</i>	U08471	-5
	U11875	-5.3
<b>Signaling</b>		
<i>Pix-<math>\alpha</math>; cool-2 (KIAA0006)</i>	D25304	-4.5
<i>ARHB/RhoB</i>	M12174	-4.5
<i>TNFSF10; TRAIL</i>	U37516	-6.6
<b>Ca<sup>2+</sup> binding</b>		
<i>ANXII</i>	L19605	-4.3
<i>S100A4</i>	M80563	-4.8
<i>ANXI</i>	X05908	-4.8
<b>Other</b>		
<i>Proteolipid protein 2</i>	L09604	-4.9
<i>Protein phosphatase 1, <math>\alpha</math> catalytic subunit</i>	HG1614-HT1614	-5
<i>TIMP2</i>	M32304	-5.1
<i>KIAA0199</i>	D83782	-5.2
<i>Lipin 3 (KIAA0249)</i>	D87436	-5.6
<i>LRMP (Jaw1)</i>	U10485	-5.8
<i>CUGBP2</i>	U69546	-6.9
<i>Clone 23933</i>	U79273	-7
<i>PECAM1</i>	L34657	-8
<i>Delta sleep-inducing peptide</i>	Z50781	-8.7
<i>DiGeorge synd. critical region gene 2 (KIAA0163)</i>	D79985	-9
<i>SELPLG; CD162; selectin P ligand</i>	U25956	-32

tion. Nevertheless, the reliance upon DNA microarrays alone affords insight only into the transcriptional response without corroboration at the protein level. In the present study, appli-

cation of both DNA microarray and proteomics technology to our model system provides unique insight into both the cellular biology of the activated PMN and the responsiveness and reg-



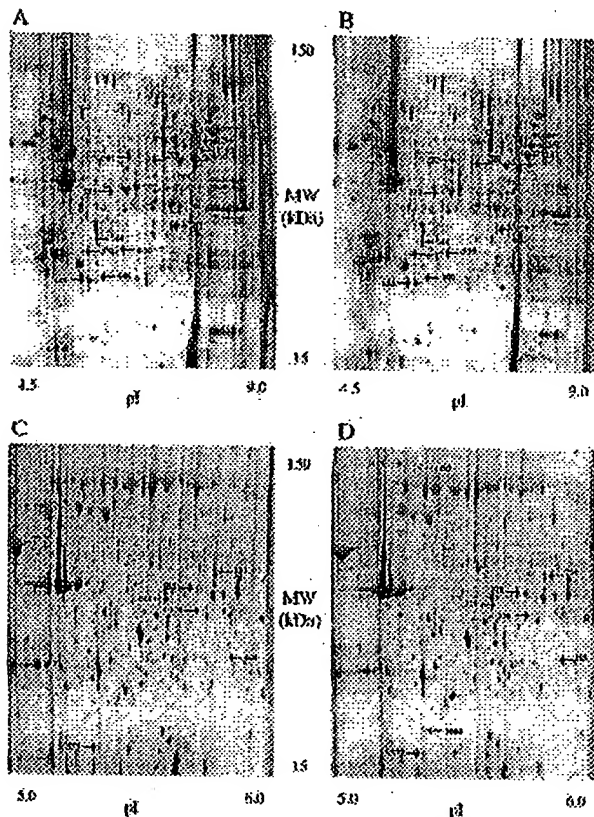


FIG. 1. Two-dimensional PAGE of LPS-exposed human PMNs. A and B, colloidal Coomassie Blue-stained pH 3.0–10.0, two-dimensional PAGE gels (A, control; B, LPS-exposed) with up-regulated (solid arrows) and down-regulated (hatched arrows) proteins indicated. These results are representative of six separate experiments. C and D, colloidal Coomassie Blue-stained pH 5.0–6.0, two-dimensional PAGE gels (C, control; D, LPS-exposed) with up-regulated (solid arrows), new (solid arrow, open arrowhead), and down-regulated (hatched arrows) proteins indicated. LPS-exposed PMNs from three blood donors were pooled.

ulation of its transcriptional and translational machinery. As will be discussed below, our study identifies, in particular, novel aspects of the LPS-stimulated PMN transcriptional regulation, activity in the innate immune response, signaling, cytoskeletal reorganization, and priming for granule release.

In the present study, the increase in NF- $\kappa$ B transcript abundance (Table I) detected by the microarrays corroborates the findings of other studies of PMNs and monocytes (40) and indicates a mechanism for the responsiveness and scope of the PMN transcriptional machinery following LPS exposure. NF- $\kappa$ B, recently described to be activated by LPS through the TLR/MyD88/interleukin-1 receptor-associated kinase pathway (1, 4), is the only transcriptional complex reported to be induced by LPS in the PMN. However, because the transcriptional NF- $\kappa$ B complex has been implicated in the regulation of only a portion of the genes induced by LPS in this study (data not shown), the importance of alternative transcriptional regulators in the PMN is clear. Of interest, several other known and putative transcriptional regulators with less well defined functions were also up-regulated in the present study, including *PLAGL2*, a putative zinc-finger protein, *XBP-1*, *MTF-1*, *Ets-2*, *B-ATF*, and *DIF-2*. On the other hand, LPS-down-regulated genes include *ATF-2* (a known target of p38), *NFATC4*, *TOB-1*, *NF-E2*, *MXI-1*, and *LYL-1*. Although the exact role of these gene products in regulating cell function is unknown,

these data indicate that the range of transcriptional responses in the LPS-stimulated PMN is much broader than previously suggested and that the signaling capabilities of the PMN in the immune response are thereby likely extended in scope and specificity.

As expected from the literature, the genes for several cytokines and chemokines, including *IL-1 $\beta$* , *IL-6*, and *MIP-1 $\beta$* , were found to be up-regulated (Table I). On the other hand, the notable absence of up-regulated cytokines in the proteomics experiments reflects their removal in the post-LPS incubation wash performed prior to lysis for two-dimensional-PAGE. Up-regulation of these inflammatory mediators is well documented in PMNs exposed to LPS and in animal models of LPS-induced sepsis syndrome and acute respiratory distress syndrome, a PMN-mediated illness (41, 42). Several genes in this family were up-regulated that have not, to our knowledge, been described in LPS-stimulated cells, including *MCP-1*, *GRO3*, *IL-10RA*, and *HM74*, an orphan G protein-coupled receptor with homology to chemokine receptors. The down-regulation of *TNFSF10*, *lymphotoxin b receptor*, and *TNFAIP1* were also observed. The modulation of genes involved in cytokine signaling, including the adapter molecules *TRAF1* (LPS and TNF receptor signaling) and *TNFAIP1* (TNF receptor signaling) and several kinases and phosphatases, may indicate a change in cytokine responsiveness after LPS treatment. Relevant in this regard from the proteomics data are: 1) the up-regulation of protein phosphatase 1, which has been shown to regulate PMN NADPH oxidase activation and translocation (43, 44) and to regulate LPS-induced NF- $\kappa$ B activation (45); 2) the down-regulation of Rho-GAP1, which has been shown to regulate NADPH oxidase activity in the PMN (46); and 3) the up-regulation of  $PO_4$ -stathmin (Table IV), a phosphoprotein postulated to function as a relay and integrator of multiple signal transduction pathways (34). Several noncytokine, nonchemokine genes involved in the immune response were also up-regulated, including the complement pathway members *C3*, *C3AR1*, and *PFC*; the protease inhibitors *ELANH2* (elastase inhibitor), *SLPI*, *PI-3*, and *PI-9*; and the acute phase protein *orosomucoid*. LPS regulation of *C3AR1* and *orosomucoid* expression have not previously been reported. In the proteomics experiments, the down-regulation of ficolin-1 (Table III), a collectin-like cell surface protein reported to activate the complement system and to mediate adhesion and phagocytosis in monocytes but not previously reported in granulocytes (47), may represent negative modulation of the innate immune response. The finding that genes other than cytokines and chemokines are regulated by the PMN in response to LPS indicates that the PMN plays a more sophisticated role in host-defense and immunity than previously thought.

Treatment of the PMN with LPS lead to the induction of a set of genes associated with the anti-viral Type I interferons, IFN $\alpha/\beta$ . This induction occurs independently of the release of IFN or another unidentified soluble factor.<sup>2</sup> Furthermore, the set of genes expressed is smaller than that induced by IFN $\alpha/\beta$ , as described by Der *et al.* (12). This may be due to differences in the scope of the signaling systems activated by LPS and IFN $\alpha/\beta$ , or the time course of analysis of genes in the LPS-stimulated PMN. The implication that LPS treatment of PMN allows PMN to express anti-viral activity is currently being tested. Of interest was the finding that induction of interferon-stimulated genes was blocked by pretreatment of PMNs with SB203580. Work from our laboratory has indicated that signal transducers and activators of transcription activation does not occur in response to LPS in PMNs.<sup>2</sup> In addition, interferon-

<sup>2</sup> K. C. Malcolm and G. S. Worthen, manuscript in preparation.

TABLE III  
Analysis of pH 3.0–10.0 two-dimensional PAGE gels

Mean change (fold) in expression level among six PMN donors is reported. The change in expression for the proteins listed was statistically significant ( $p < 0.05$ ) as measured by a two-tailed Student's *t* test.

Identification [spot no.]	Swiss-Prot no.	Estimated $M_R$ /pI	Theoretical $M_R$ /pI	Peptides matched/ submitted	Protein covered	Mean change
				%	%	-fold
<b>Up-regulated</b>						
Proteasome $\beta$ chain [646]	P28070	27/5.7	29.2/5.72	9/12 (75%)	36%	1.51
Annexin III [550]	P12429	31/5.7	36.4/5.6	14/18 (78%)	42%	1.37
Actin fragment [544] <sup>a</sup>	P02570	32/5.5	(41.7/5.29)	13/15 (87%)	(34%)	1.74
Actin fragment [591] <sup>a</sup>	P02570	30/5.4	(41.7/5.29)	14/18 (78%)	(29%)	1.60
$\alpha$ -Enolase [380]	P06733	41/5.7	47.2/7.01	9/10 (90%)	24%	1.65
Rab-GDP dissociation inhibitor $\beta$ [289]	P50395	50/6.1	50.7/6.11	10/11 (91%)	25%	1.24
Glutathione S-transferase P [648]	P09211	23/5.5	23.4/5.43	6/8 (75%)	41%	1.54
Pre-B-cell colony enhancing factor [1152]	P43490	53/7.0	55.5/6.69	12/16 (75%)	25%	1.29
<b>Down-regulated</b>						
Adenylyl cyclase-associated protein 1 [256]	Q01518	55/7.3	51.7/8.07	16/22 (73%)	34%	0.53
Rho-GAP1 [283]	Q07960	50/5.8	50.4/5.85	7/9 (78%)	22%	0.67
Ficolin 1 [511]	O00602	33/6.5	35/6.39	10/12 (83%)	25%	0.74

<sup>a</sup> The theoretical pI and  $M_R$  of native actin are indicated. Protein coverage indicates coverage of native actin.

TABLE IV  
Analysis of pH 5.0–6.0 two-dimensional PAGE gels

Results are from pooled samples for control ( $n=3$ ) and LPS-exposed ( $n=3$ ) PMNs from human donors. Expression of the reported proteins was altered >1.5-fold following LPS exposure in two repeat experiments. "New" designates proteins seen in the LPS gel in two repeat experiments but not detectable in the corresponding control gels.

Identification [spot no.]	Swiss-Prot no.	Estimated $M_R$ /pI	Theoretical $M_R$ /pI	Peptides matched/ submitted	Protein covered	Change
				%	%	-fold
<b>Up-regulated</b>						
Protein-tyrosine kinase 9-like [468]	Q9Y3F5 <sup>a</sup>	34/5.81	39.5/6.37	10/14 (71%)	34%	1.8
Protein phosphatase 1, catalytic subunit, $\beta$ isoform [378]	P37140	38/5.73	37.2/5.84	7/10 (70%)	22%	2.0
PO <sub>4</sub> -stathmin [577]	P16949 <sup>b</sup>	18/5.36	17.3/5.76	9/12 (75%)	42%	2.1 <sup>c</sup>
Nonmuscle myosin heavy chain [1102]	189036 <sup>c</sup>	145/5.32	145/5.23	20/21 (95%)	17%	New
Putative PO <sub>4</sub> -nonmuscle myosin heavy chain [1101] <sup>d</sup>	189036 <sup>b,c</sup>	145/5.29	145/5.23	14/16 (87%)	13%	New
Leukocyte elastase inhibitor [318]	P30740	42/5.71	42.7/5.9	9/13 (69%)	22%	2.4
Grancalcin [1004]	P28676	24/5.36	24.0/5.02	7/10 (70%)	31%	New
<b>Down-regulated</b>						
Adenosylhomocysteinase [324]	P23526	48/5.82	47.7/6.04	7/9 (78%)	14%	0.4
PEST phosphatase interacting protein homolog [234] <sup>e</sup>	4100162 <sup>f</sup>	48/5.30	47.6/5.35	11/13 (85%)	30%	0.5

<sup>a</sup> TrEMBL accession number.

<sup>b</sup> Accession number and theoretical pI and  $M_R$  for the unmodified protein are indicated.

<sup>c</sup> NCBI accession number.

<sup>d</sup> See text for explanation.

<sup>e</sup> Among three experiments, the ratio of PO<sub>4</sub>-stathmin expression increase, following LPS exposure in the presence of SB203580 divided by that in the absence of SB203580, was 0.93.

<sup>f</sup> Genpept accession number.

<sup>g</sup> This search was performed using average masses measured by linear mode MALDI-TOF MS.

TABLE V  
Analysis of pH 5.5–6.7 two-dimensional PAGE gels

Results are from pooled samples for control ( $n=3$ ) and LPS-exposed ( $n=3$ ) PMNs from human donors. Expression of the reported proteins was altered >1.5-fold following LPS exposure in two repeat experiments.

Identification [spot no.]	Swiss-Prot no.	Estimated $M_R$ /pI	Theoretical $M_R$ /pI	Peptides matched/ submitted	Protein covered	Change
				%	%	-fold
<b>Up-regulated</b>						
Transaldolase [476]	P37837	38/5.95	37.5/6.36	13/17 (76%)	33%	2.5
Isocitrate dehydrogenase [431]	O75874	46/6.25	46.7/6.35	7/7 (100%)	13%	2.3
Moesin [201]	P26038	61/6.09	67.8/6.07	11/13 (85%)	17%	2.1
$\alpha$ -Enolase [459]	P06733	43/5.64	47.2/7.01	7/10 (70%)	17%	3.8
<b>Down-regulated</b>						
Calponin H2 [240]	Q99439	34/6.65	33.7/6.94	10/11 (90%)	27%	0.5

regulatory factor 3, a known regulator of interferon-stimulated gene transcription, is not a direct target of p38 kinase.<sup>2</sup> Therefore, gene expression analysis of LPS-stimulated PMNs has uncovered a previously uncharacterized signal transduction system that is sensitive to inhibition of p38 MAPK.

Knowledge of the genes down-regulated by LPS permits the

development of further hypotheses addressing PMN function in the face of infection. Strikingly, several down-regulated genes and gene products are structural in nature (*e.g.* paxillin, actinin, calponin H2) (Tables II and V). A known consequence to the PMN of LPS exposure is decreased motility (48). Up-regulation of genes for adhesion molecules (*ICAM-1*, *CD44*, *AL-*

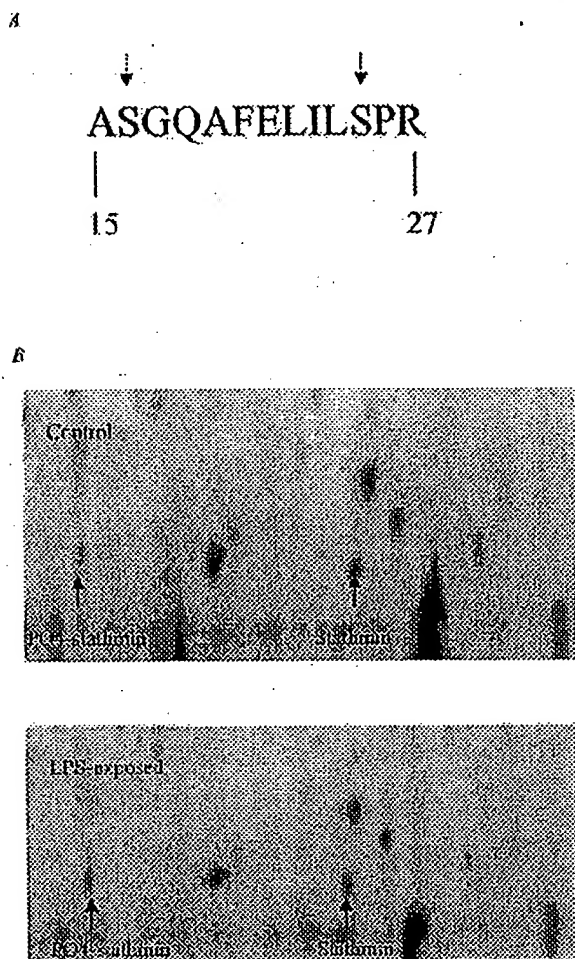


FIG. 2. A, the predicted sequence of the tryptic phosphopeptide in  $\text{PO}_4$ -stathmin (1468.72 Da). The peptide mass measured by MALDI-TOF MS and the predicted mass differed by 14 ppm. As indicated, two alternate phosphorylation sites are possible: serine 16 and serine 25. B,  $\text{PO}_4$ -stathmin and stathmin were identified on the control and LPS-exposed pH 5.0–6.0 gels. Consistent with phosphorylation, the  $\text{PO}_4$ -stathmin spot was distinguished by a peptide of mass 1468.72 Da (i.e. 80 Da greater than the peptide of 1388.72 Da seen in the stathmin spot). Assuming that no other multiply phosphorylated stathmin species have escaped detection, analysis of the integrated intensities of the  $\text{PO}_4$ -stathmin and stathmin spots indicates that the percentage of the  $\text{PO}_4$  form of total cellular stathmin has increased from 11% to 38% with LPS stimulation. The decrease in integrated intensity for stathmin was equal in amount to the increase in  $\text{PO}_4$ -stathmin following LPS exposure.

CAM, and TSG-6), and down-regulation of genes for structural proteins, indicates a genetic basis for this observation. Down-regulation of two genes implicated in cytoskeletal regulation, *Fix- $\alpha$*  and *RhoB*, was also observed. The calcium-binding protein S100A4, down-regulated in LPS-treated PMNs (Table II), has been implicated in cell motility and metastasis (49). Decreased motility may be beneficial in sustaining the inflammatory response at sites of infection. In addition, LPS treatment results in an inhibition of apoptosis (50). Therefore, the longer residence time of the PMN at sites of infection is consistent with the long term genetically coded changes seen in these gene-profiling experiments and indicates that the changes in gene expression are functionally relevant to host defense and immunity.

By providing information on post-translational modification, the proteomics data may provide further insights into the cy-

TABLE VI  
Effect of SB203580 on LPS-stimulated gene expression  
Genes are reported for which the SB203580/control expression ratio is  $\leq 0.60$ .

Gene name	-fold change ratio (SB203580/control)	Change in absence of SB203580
		-fold
<i>ISG15</i>	0.09	22.5
<i>HCR</i>	0.38	20.8
<i>Mx-1</i>	0	19.4
<i>IFI56</i>	0	12.1
<i>PI-9</i>	0.57	9.5
<i>Ets-2</i>	0.59	7.4
<i>IL-6</i>	0.45	6.3
<i>Rel</i>	0.50	6.2
<i>LIMS1</i>	0.58	6.1
<i>C3AR1</i>	0.49	6.1
<i>INDO</i>	0.35	5.2
<i>KIAA0105</i>	0.41	5.1
<i>SNAP23</i>	0.58	6.0
<i>SLP1</i>	0.58	4.7
<i>ELNAH2</i>	0.49	4.6
<i>HM-74</i>	0.57	3.8
<i>PKR</i>	0	3.7
<i>MAD</i>	0.21	3.6
<i>IFIT4</i>	0.12	3.6
<i>Glycerol kinase</i>	0	3.6
<i>IFI54</i>	0	3.5
<i>IFI58</i>	0.39	3.5
<i>IPF35</i>	0.46	3.0

TABLE VII  
Effect of SB203580 on LPS-stimulated protein expression

Protein name	-fold change ratio (SB203580/control)	Change in absence of SB203580
		-fold
<i>Up-regulated</i>		
Proteasome $\beta$ chain	0.8	1.51
Annexin III	0.6	1.37
Actin fragment [544]	0.8	1.74
Actin fragment [591]	0.8	1.60
$\alpha$ -Enolase	0.6	1.65
Rab-GDP dissociation inhibitor $\beta$	1.1	1.24
Glutathione S-transferase P	1.2	1.54
Pre-B-cell colony enhancing factor	1.2	1.29
<i>Down-regulated</i>		
Adenylyl cyclase-associated protein 1	1.3	0.53
Rho-GAP1	0.8	0.67
Ficolin 1	1.0	0.74

toskeletal remodeling effects of LPS upon the PMN. We contend that the actin fragments identified (Table III) are unlikely to represent technical artifacts. Rather, their specificity (identical molecular weight/pI among different experiments), statistically significant up-regulation by LPS, as well as the use of a lysis buffer containing chaotropes and multiple protease inhibitors argue instead that these fragments are physiologic consequences of LPS exposure in the human PMN. More specifically, the up-regulation of these fragments following LPS exposure (Table III) suggests that LPS may activate an actin-cleaving enzyme, which, in turn, remodels the cytoskeleton. Intriguing in this vein, calpain has recently been reported to play an important role in cell migration and cytoskeletal organization of fibroblasts (51). The possibilities that LPS may induce calpain activation and that calpain activation may regulate cytoskeletal reorganization and motility are currently under investigation. An alternative possibility is that actin cleavage is a marker of neutrophil apoptosis (52).

Other LPS-regulated proteins may play important roles in cytoskeletal reorganization. The up-regulation of protein-tyrosine kinase 9-like (A6-related protein) may modulate LPS-

TABLE VIII  
LPS-regulated proteins for which a probe was present on the  
Affymetrix chip

A comparison of corresponding protein and mRNA transcript changes following LPS exposure is shown.

Protein	Protein change	mRNA change
<i>Up-regulated</i>		
Proteasome $\beta$ chain	1.5	1.9 $\uparrow$
Leukocyte elastase inhibitor	2.4	4.6 $\uparrow$
Rab-GDI $\beta$	1.24	NC <sup>a</sup>
Grancalcin	New	NC
Transaldolase	2.5	NC
Moesin	2.1	NC
Nonmuscle myosin heavy chain	New	NC
Glutathione <i>S</i> -transferase P	1.54	Absent
Pre-B cell enhancing factor	1.29	Absent
Isocitrate dehydrogenase	2.3	Absent
PO <sub>4</sub> -stathmin	2.1	Absent (stathmin)
Protein phosphatase 1, $\beta$ catalytic subunit	2	Absent
Annexin III	3.1	3.1 $\downarrow$
<i>Down-regulated</i>		
Adenyl cyclase-associated protein 1	1.9	2.1 $\downarrow$
Rho-GAP 1	1.5	2.7 $\downarrow$
Ficolin 1	1.4	1.7 $\downarrow$
Adenosylhomocysteinase	2.5	Absent
Calponin H2	2	NC

<sup>a</sup> NC, no measureable change.

induced actin polymerization, because it bears a high degree of homology to twinfilin (A6), an actin monomer-binding protein that localizes to sites of rapid filament assembly in cells and is believed to regulate actin filament turnover (53). In turn, LPS-induced down-regulation of Rho-GTPase activating protein 1 (Table III) may regulate twinfilin (and protein-tyrosine kinase 9-like) activity, because twinfilin has been shown to colocalize with Rac1 and Cdc42 and to be regulated by active Rac1 in NIH 3T3 cells (53). Activation of Rho proteins may be facilitated by LPS up-regulation of moesin (Table V), because moesin reportedly induces the dissociation of Rho from GDI (54). Rac1 may, in turn, promote activation of the actin filament-nucleating Arp2/3 complex through interactions with WASP (Wiskott-Aldrich syndrome protein) family proteins (55) and, interestingly, is postulated to regulate the dynamics of both the actin and microtubule cytoskeletons via phosphorylation of stathmin (Table IV) (56). Calponin H2 is an actin-binding protein not previously reported in PMNs that is postulated to play a role in cytoskeletal organization (57). Its down-regulation by LPS (Table V) likely modulates LPS-induced cytoskeletal reorganization. The up-regulation of nonmuscle myosin heavy chain and a putative phosphorylated form of myosin heavy chain (putative protein kinase C substrate by prediction rules) in the LPS-exposed PMN (Table IV) is of uncertain significance; myosin has been implicated in multiple functions in the PMN, including locomotion, fluid pinocytosis, and phagocytosis (58). Of interest, however, S100A4 (down-regulated, Table II) has been reported to regulate cytoskeletal dynamics by inhibiting protein kinase C-mediated phosphorylation of nonmuscle myosin heavy chain (59).

LPS induction of stathmin phosphorylation (Table IV and Fig. 2) may represent another mechanism by which the cytoskeleton is remodeled. Stathmin is a phosphoprotein reportedly involved in both signal transduction and in regulation of the microtubulin filament network; furthermore, phosphorylation of stathmin has been reported to modulate its tubulin-binding avidity (60). Inferences can be made about both the phosphorylation site on PO<sub>4</sub>-stathmin and the responsible kinase induced by LPS. Four phosphorylation sites in stathmin have been well described: Ser<sup>16</sup>, Ser<sup>26</sup>, Ser<sup>38</sup>, and Ser<sup>63</sup> (32, 33).

Ser<sup>16</sup> has been reported as a substrate for Ca<sup>2+</sup>/calmodulin (CaM)-dependent kinases (32), and Ser<sup>26</sup> as primarily a substrate for p38 and ERK (33), with p34<sup>cdc2</sup> also active but bearing a 5-fold preference for Ser<sup>36</sup> (34). As stated above, the phosphopeptide identified in PO<sub>4</sub>-stathmin, extending from residues 15 to 27 (1468.7 Da), is consistent with phosphorylation of either Ser<sup>16</sup> or Ser<sup>26</sup> (Fig. 2). Although both p38 $\delta$  and p38 $\alpha$  MAPK isoforms are expressed in the human PMN, LPS has been shown to selectively activate the p38 $\alpha$  isoform in human PMNs (9). The p38 $\alpha$  isoform, however, has been shown to be relatively inactive at Ser<sup>26</sup>; in fact, p38 $\delta$  is ~100-fold more active at Ser<sup>26</sup>, and selective p38 $\alpha$  inhibitors do not inhibit the stress-activated phosphorylation of stathmin in 293 cells (33). Further support for the lack of involvement of p38 signaling in phosphorylation of stathmin in our system is the apparent lack of effect of SB203580 (a selective p38 $\alpha$  and p38 $\beta$  inhibitor) on LPS-induced expression of PO<sub>4</sub>-stathmin (Table IV). Because p34<sup>cdc2</sup> is relatively inactive at Ser<sup>26</sup> (34), we conclude that the phosphorylation site is likely to be Ser<sup>16</sup>, a reported substrate of CaM-dependent kinase. Although CaM kinases have previously been implicated in gene activation in LPS-exposed myelomonocytic HD11 cells (61), stathmin signaling has not, to our knowledge, been previously reported in either PMNs or lipopolysaccharide signal transduction.

Cytoskeletal reorganization, a well-described regulator of granule release (62), may underlie LPS-induced priming for PMN granule release, but several LPS-regulated proteins may provide more specific clues. LPS exposure led to increased levels of grancalcin, a calcium-binding protein previously detected in PMNs and shown to translocate to granules and plasma membrane in the presence of physiologic concentrations of calcium (63). Similarly, annexin III, a calcium-binding protein highly expressed in PMN granule membranes and implicated in calcium-mediated secretion (64) and in granule fusion (65), was also found to be up-regulated. Exocytosis of granule contents may also be facilitated by LPS up-regulation of Rab-GDP dissociation inhibitor (Table III), which has been proposed to recycle Rab after vesicle fusion by extracting it from the membrane and loading it onto newly formed transport intermediates (66).

Parallel use of DNA microarrays and proteomics affords a powerful strategy for comparison of corresponding mRNA transcripts and proteins, thereby affording new insight into the mechanisms by which the cell regulates its signaling responses to the external environment. Of interest, a poor correlation was found between corresponding transcripts and proteins (Table VIII), as reported in other systems (17, 18). The finding in some cases of unchanged transcript abundance in the face of regulated protein levels indicates post-transcriptional modulation following LPS exposure. The finding of undetected transcripts in the face of regulated levels of the corresponding proteins may indicate previous transcription of these genes in an earlier state of the myeloid maturation of the PMN, producing stable protein species that have undergone post-translational alteration following LPS exposure. The use of SB203580, a p38 inhibitor, adds further insights into the mechanisms of LPS regulation. At the level of mRNA expression, SB203580 inhibited 23% of LPS-stimulated genes by  $\geq 40\%$  and 11% of genes by  $\geq 60\%$ ; therefore, p38 plays a specific role in gene regulation in the PMN. In particular, proteasome  $\beta$  chain was up-regulated at both the mRNA transcript and protein level (Table VIII), with no notable effect of SB203580 on expression at either level, consistent with a non-p38-mediated pathway of primary transcriptional up-regulation induced by LPS. Similarly, CAP1, Rho-GAP1, and ficolin 1 were down-regulated at both the mRNA transcript and protein level (Table VIII), with

no notable effect of SB203580, consistent with a non-p38-mediated pathway of primary transcriptional down-regulation. Interestingly, annexin III was down-regulated at the transcript level and up-regulated at the protein level, with an inhibitory effect of SB203580 seen only at the protein level (Table VII), consistent with a p38-mediated post-transcriptional up-regulation induced by LPS.

Limitations of the present study should be noted. Gene expression analysis by cDNA microarrays does not distinguish between transcriptional regulation and mRNA stabilization; similarly, two-dimensional PAGE proteomics by itself does not distinguish among transcriptional, translational, or post-translational regulation of protein abundance. Transcript detection by microarray technology is limited to the probes included; protein identification by two-dimensional PAGE proteomics is limited to well-resolved regions of the gel, may perform less well with hydrophobic and high molecular weight proteins, and tends to select for more abundant protein species (30). Harvesting of the LPS-incubated PMNs at 4 h may have prevented detection of earlier, transient changes and may have thereby introduced artifactual transcript-protein discordance. Furthermore, the post-LPS incubation, pre-two-dimensional PAGE cell washes would be expected to remove secreted proteins from further analysis, with uncertain effects on detected protein abundance depending on such factors as the degree of *de novo* synthesis and extent of degranulation/exocytosis. Because protein binding of Coomassie Blue has a limited dynamic range and is typically not linear throughout the range of detection, image analysis of Coomassie Blue-stained protein spots should be considered semi-quantitative. For some protein spots, the apparent magnitude of regulation by LPS may have been blunted by the spot approaching staining saturation in the control gel. By limiting our analysis to those protein spots common to all twelve pH 3.0–10.0 two-dimensional gels, we likely excluded some LPS-regulated proteins that happened to be either poorly resolved on a subset of the gels or unmatched by the image analysis software. By further limiting the analysis to those matched spots on the pH 3.0–10.0 gels for which a two-tailed *t* test demonstrated  $p < 0.05$ , the list of regulated proteins was likely also limited by statistical power. In addition to those regulated proteins listed in Table III, three others were up-regulated and three down-regulated with  $p < 0.09$  (data not shown).

Limiting our reported results to those changes that met statistical significance among the donors carries further important implications. We have encountered a two order of magnitude range of response in unselected donor LPS-induced PMN functions, such as TNF- $\alpha$  and superoxide anion release (data not shown). The sources of this physiologic heterogeneity remain uncertain but may possibly include such factors as natural mutations of the LPS receptor component, TLR4 (67). By selecting for LPS effects common to all donors, we may not have characterized the range of genomic and proteomic heterogeneity present in the population and thereby may have focused on only a narrow portion of a broader biological response to LPS. We contend that this reductionist approach is valid because it would be expected to enrich for biologically integral responses of the PMN to LPS. Nevertheless, correlation of genomic and proteomic profiles with functional phenotypes of the PMN may bear important diagnostic and therapeutic implications and will be pursued in future studies.

Widespread regulation of numerous noncytokine/chemokine genes and proteins in the LPS-stimulated human PMN is a novel finding. These data indicate that, despite a narrow scope of gene expression in the nonstimulated state, the terminally differentiated, short-lived PMN likely plays a role in the innate

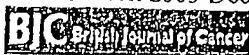
immune response that is far more sophisticated and dynamic than the simple release of preformed inflammatory mediators. Although gene expression appears to be an important mechanism by which PMNs respond acutely to infection, mRNA transcript/protein concordance is limited, and post-transcriptional (and post-translational) modifications also play an important role. The alteration of multiple transcriptional regulators, G-protein regulators, PO<sub>4</sub>-stathmin, and protein phosphatase 1 indicates that one of the responses to LPS exposure is to modify subsequent signaling events by bacterial components or by other cytokines and chemokines. Finally, the finding that p38 MAPK mediates LPS regulation of a limited subset of transcripts and proteins underlines the continuing need to define signal transduction cascades in the neutrophil.

**Acknowledgments**—We thank the members of the Affymetrix core laboratory, University of Colorado Health Sciences Center, as well as Benjamin Perryman, Steve Helmke, and Jennifer Lynch of the Cardiology Division, University of Colorado Health Sciences Center for assistance with two-dimensional PAGE.

#### REFERENCES

- Bowie, A., and O'Neill, L. A. (2000) *J. Leukocyte Biol.* 67, 508–514
- Chow, J. C., Young, D. W., Golenbock, D. T., Christ, W. J., and Gusovsky, F. (1999) *J. Biol. Chem.* 274, 10689–10692
- Kirschning, C. J., Wesche, H., Merrill Ayres, T., and Rothe, M. (1998) *J. Exp. Med.* 188, 2091–2097
- Muzio, M., Polentarutti, N., Bosio, D., Praladani, M. K., and Mantovani, A. (2000) *J. Leukocyte Biol.* 67, 450–456
- Yang, R. B., Mark, M. R., Gray, A., Huang, A., Xie, M. H., Zhang, M., Goddard, A., Wood, W. L., Gurney, A. L., and Godowski, P. J. (1998) *Nature* 395, 284–288
- Parsons, P. E., Worthen, G. S., Moore, E. E., Tate, R. M., and Henson, P. M. (1989) *Am. Rev. Respir. Dis.* 140, 294–301
- Nick, J. A., Avdi, N. J., Young, S. K., McDonald, P. P., Billstrom, M. A., Henson, P. M., Johnson, G. L., and Worthen, G. S. (1999) *Chest* 116, 543–555
- Nick, J. A., Young, S. K., Brown, K. K., Avdi, N. J., Arndt, P. G., Suratt, B. T., James, M. S., Henson, P. M., and Worthen, G. S. (2000) *J. Immunol.* 164, 2151–2159
- Nick, J. A., Avdi, N. J., Young, S. K., Lehman, L. A., McDonald, P. P., Frasch, S. C., Billstrom, M. A., Henson, P. M., Johnson, G. L., and Worthen, G. S. (1999) *J. Clin. Invest.* 103, 851–858
- Cassatella, M. A. (1995) *Immunol. Today* 16, 21–26
- McDonald, P. P., Bald, A., and Cassatella, M. A. (1997) *Blood* 89, 3421–3433
- Der, S. D., Zhou, A., Williams, B. R., and Silverman, R. H. (1998) *Proc. Natl. Acad. Sci. U. S. A.* 95, 15623–15628
- Karpf, A. R., Peterson, P. W., Rawlins, J. T., Dalley, B. K., Yang, Q., Albertsen, H., and Jones, D. A. (1999) *Proc. Natl. Acad. Sci. U. S. A.* 96, 14007–14012
- Iyer, V. R., Eisen, M. B., Ross, D. T., Schuler, G., Moore, T., Lee, J. C., Trent, J. M., Staudt, L. M., Hudson, J., Jr., Boguski, M. S., Lashkari, D., Shalon, D., Botstein, D., and Brown, P. O. (1999) *Science* 283, 83–87
- Han, J., Jiang, Y., Li, Z., Kravchenko, V. V., and Ulevitch, R. J. (1997) *Nature* 386, 290–299
- Lee, J. C., Laydon, J. T., McDonnell, P. C., Gallagher, T. F., Kumar, S., Green, D., McNulty, D., Blumenthal, M. J., Heys, J. R., Landvatter, S. W. et al. (1994) *Nature* 372, 739–746
- Andersson, L., and Seilhamer, J. (1997) *Electrophoresis* 18, 533–537
- Gygi, S. P., Rochon, Y., Franz, B. R., and Aebersold, R. (1999) *Mol. Cell. Biol.* 19, 1720–1730
- Lewis, T. S., Hunt, J. B., Aveline, L. D., Jonscher, K. R., Louie, D. P., Yeh, J. M., Nahraini, T. S., Resing, K. A., and Ahn, N. G. (2000) *Mol. Cell* 6, 1343–1354
- Soskic, V., Gorlach, M., Poznanovic, S., Boehmer, F. D., and Godovac-Zimmermann, J. (1999) *Biochemistry* 38, 1757–1764
- Triantafyllou, K., Triantafyllou, M., and Dedrick, R. L. (2001) *Nat. Immunol.* 2, 338–345
- Nick, J. A., Avdi, N. J., Young, S. K., Knall, C., Gerwins, P., Johnson, G. L., and Worthen, G. S. (1997) *J. Clin. Invest.* 99, 875–886
- Rabilloud, T., Valette, C., and Lawrence, J. J. (1994) *Electrophoresis* 15, 1552–1558
- Bradford, M. M. (1976) *Anal. Biochem.* 72, 248–254
- Neuhoff, V., Arnold, N., Taube, D., and Ehrhardt, W. (1988) *Electrophoresis* 9, 255–262
- Arnott, D., O'Connell, K. L., King, K. L., and Stults, J. T. (1998) *Anal. Biochem.* 258, 1–18
- Hellman, U., Wernstedt, C., Ganev, J., and Heldin, C. H. (1995) *Anal. Biochem.* 224, 451–455
- Yan, J. X., Sanchez, J. C., Rouge, V., Williams, K. L., and Hochstrasser, D. F. (1999) *Electrophoresis* 20, 723–726
- Mardian, J. K., and Isenberg, I. (1978) *Anal. Biochem.* 81, 1–12
- Corthals, G. L., Wasinger, V. C., Hochstrasser, D. F., and Sanchez, J. C. (2000) *Electrophoresis* 21, 1104–1115
- Blom, N., Gammeltoft, S., and Brumak, S. (1999) *J. Mol. Biol.* 294, 1351–1362
- Le Gouvello, S., Manceau, V., and Sobel, A. (1998) *J. Immunol.* 161, 1113–1122
- Parker, C. G., Hunt, J., Diener, K., McGinley, M., Soriano, B., Keesler, G. A., Bray, J., Yao, Z., Wang, X. S., Kohno, T., and Lichenstein, H. S. (1998)

- Biochem. Biophys. Res. Commun.* 249, 791-796
34. Marklund, U., Brattsand, G., Shingler, V., and Gullberg, M. (1993) *J. Biol. Chem.* 268, 15039-15047
  35. Cohen, P., Bouaboula, M., Bellis, M., Baron, V., Jbilo, O., Poinet-Chazel, C., Galiegua, S., Hadjibi, E. H., and Casellas, P. (2000) *J. Biol. Chem.* 275, 11181-11190
  36. Yoshimura, A., Lien, E., Ingalls, R. R., Tuomanen, E., Dziarski, R., and Golenbock, D. (1999) *J. Immunol.* 163, 1-5
  37. Lien, E., Sellati, T. J., Yoshimura, A., Flo, T. H., Rawadi, G., Finberg, R. W., Carroll, J. D., Espevik, T., Ingalls, R. R., Radolf, J. D., and Golenbock, D. T. (1999) *J. Biol. Chem.* 274, 33419-33425
  38. Underhill, D. M., Ozinsky, A., Hajjar, A. M., Stevens, A., Wilson, C. B., Bassetti, M., and Aderem, A. (1999) *Nature* 401, 811-815
  39. Brightbill, H. D., Libraty, D. H., Krutzik, S. R., Yang, R. B., Belisle, J. T., Bleharski, J. R., Maitland, M., Norgard, M. V., Plevy, S. E., Smale, S. T., Brennan, P. J., Bloom, B. R., Godowski, P. J., and Modlin, R. L. (1999) *Science* 285, 732-736
  40. de Wit, H., Dokter, W. H., Koopmans, S. B., Lummen, C., van der Leij, M., Smit, J. W., and Vellenga, E. (1998) *Leukemia* 12, 363-370
  41. Johnston, C. J., Finkelstein, J. N., Gelein, R., and Oberdorster, G. (1998) *Toxicol. Sci.* 48, 300-307
  42. Ulich, T. R., Watson, L. R., Yin, S. M., Guo, K. Z., Wang, P., Thang, H., and del Castillo, J. (1991) *Am. J. Pathol.* 138, 1485-1495
  43. Karlsson, A., Nixon, J. B., and McPhail, L. C. (2000) *J. Leukocyte Biol.* 67, 396-404
  44. Dorseuil, O., Quinn, M. T., and Bokoch, G. M. (1995) *J. Leukocyte Biol.* 58, 108-113
  45. Pahan, K., Sheikh, F. G., Nambodiri, A. M., and Singh, I. (1998) *J. Biol. Chem.* 273, 12219-12226
  46. Geiszt, M., Dagher, M. C., Molnar, G., Havasi, A., Faure, J., Paclet, M. H., Morel, F., and Ligeti, E. (2001) *Biochem. J.* 355, 851-858
  47. Teh, C., Le, Y., Lee, S. H., and Lu, J. (2000) *Immunology* 101, 225-232
  48. Wagner, J. G., and Roth, R. A. (1999) *J. Leukocyte Biol.* 66, 10-24
  49. Barraclough, R. (1998) *Biochim. Biophys. Acta* 1448, 190-199
  50. Lee, A., Whyte, M. K., and Haslett, C. (1993) *J. Leukocyte Biol.* 54, 283-288
  51. Dourdin, N., Bhatt, A. K., Dutt, P., Greer, P. A., Arthur, J. S., Elee, J. S., and Huttenlocher, A. (2001) *J. Biol. Chem.* 276, 48362-48366
  52. Brown, S. B., Bailey, K., and Savill, J. (1997) *Biochem. J.* 323, 233-237
  53. Vartiainen, M., Ojala, P. J., Auvinen, P., Peranen, J., and Lappalainen, P. (2000) *Mol. Cell. Biol.* 20, 1772-1783
  54. Takahashi, K., Sasaki, T., Mammoto, A., Takashi, K., Kameyama, T., Tsukita, S., and Takai, Y. (1997) *J. Biol. Chem.* 272, 23371-23375
  55. Macheeky, L. M., and Insall, R. H. (1998) *Curr. Biol.* 8, 1347-1356
  56. Daub, H., Gevaert, K., Vandekerckhove, J., Sobel, A., and Hall, A. (2001) *J. Biol. Chem.* 276, 1677-1680
  57. Masuda, H., Tanaka, K., Takagi, M., Ohgami, K., Sakamaki, T., Shibata, N., and Takahashi, K. (1996) *J. Biochem. (Tokyo)* 120, 415-424
  58. Valerius, N. H., Stendahl, O. I., Hartwig, J. H., and Stossel, T. P. (1982) *Adv. Exp. Med. Biol.* 141, 19-28
  59. Krijanvaska, M., Tarabykina, S., Bronstein, I., Maitland, N., Lomonosov, M., Hansen, K., Georgiev, G., and Lukamidin, E. (1998) *J. Biol. Chem.* 273, 9852-9856
  60. Steinmetz, M. O., Jalinka, W., Towbin, H., Garcia-Echeverria, C., Voshol, H., Muller, D., and van Oostrum, J. (2001) *EMBO Rep.* 2, 505-510
  61. Regenhart, P., Goetha, R., and Phi-van, L. (2001) *J. Leukocyte Biol.* 69, 651-658
  62. Valentijn, K., Valentijn, J. A., and Jamieson, J. D. (1999) *Biochem. Biophys. Res. Commun.* 268, 652-661
  63. Toahan, C. G., Totty, N. F., and Segal, A. W. (1992) *Biochem. J.* 286, 549-554
  64. Rosales, J. L., and Ernst, J. D. (1997) *J. Immunol.* 159, 6195-6202
  65. Le Cabec, V., and Maridonneau-Parini, I. (1994) *Biochem. J.* 303, 481-487
  66. Gilbert, P. M., and Burd, C. G. (2001) *J. Biol. Chem.* 276, 8014-8020
  67. Arbour, N. C., Lorenz, E., Schutte, B. C., Zabner, J., Kline, J. N., Jones, M., Frees, K., Watt, J. L., and Schwartz, D. A. (2000) *Nat. Genet.* 25, 187-191



**An increased high-mobility group A2 expression level is associated with malignant phenotype in pancreatic exocrine tissue.**

Abe N, Watanabe T, Suzuki Y, Matsumoto N, Masaki T, Mori T, Sugiyama M, Chiappetta G, Fusco A, Atomi Y.

First Department of Surgery, Kyorin University School of Medicine, 6-20-2, Shinkawa, Mitaka, Tokyo 181-8611, Japan. abenbtg@kyorin-u.ac.jp

The altered form of the high-mobility group A2 (HMGA2) gene is somehow related to the generation of human benign and malignant tumours of mesenchymal origin. However, only a few data on the expression of HMGA2 in malignant tumour originating from epithelial tissue are available. In this study, we examined the HMGA2 expression level in pancreatic carcinoma, and investigated whether alterations in the HMGA2 expression level are associated with a malignant phenotype in pancreatic tissue. High-mobility group A2 mRNA and protein expression was determined in eight surgically resected specimens of non-neoplastic tissue (six specimens of normal pancreatic tissue and two of chronic pancreatitis tissue) and 27 pancreatic carcinomas by highly sensitive reverse transcriptase-polymerase chain reaction (RT-PCR) techniques and immunohistochemical staining, respectively. Reverse transcriptase-polymerase chain reaction analysis revealed the expression of the HMGA2 gene in non-neoplastic pancreatic tissue, although its expression level was significantly lower than that in carcinoma. Immunohistochemical analysis indicated that the presence of the HMGA2 gene in non-neoplastic pancreatic tissue observed in RT-PCR reflects its abundant expression in islet cells, together with its focal expression in duct epithelial cells. Intense and multifocal or diffuse HMGA2 immunoreactivity was noted in all the pancreatic carcinoma examined. A strong correlation between HMGA2 overexpression and the diagnosis of carcinoma was statistically verified. Based on these findings, we propose that an increased expression level of the HMGA2 protein is closely associated with the malignant phenotype in the pancreatic exocrine system, and accordingly, HMGA2 could serve as a potential diagnostic molecular marker for distinguishing pancreatic malignant cells from non-neoplastic pancreatic exocrine cells.

PMID: 14647145 [PubMed - indexed for MEDLINE]

# MOLECULAR BIOLOGY OF **THE CELL**

## THIRD EDITION

Bruce Alberts • Dennis Bray  
Julian Lewis • Martin Raff • Keith Roberts  
James D. Watson



Garland Publishing, Inc.  
New York & London



## GARLAND STAFF

Text Editor: Miranda Robertson  
 Managing Editor: Ruth Adams  
 Illustrator: Nigel Orme  
 Molecular Model Drawings: Kate Hesketh-Moore  
 Director of Electronic Publishing: John M-Roblin  
 Computer Specialist: Chuck Bartelt  
 Disk Preparation: Carol Winter  
 Copy Editor: Shirley M. Cobert  
 Production Editor: Douglas Goertzen  
 Production Coordinator: Perry Bessas  
 Indexer: Maija Hinkle

Bruce Alberts received his Ph.D. from Harvard University and is currently President of the National Academy of Sciences and Professor of Biochemistry and Biophysics at the University of California, San Francisco. Dennis Bray received his Ph.D. from the Massachusetts Institute of Technology and is currently a Medical Research Council Fellow in the Department of Zoology, University of Cambridge. Julian Lewis received his D.Phil. from the University of Oxford and is currently a Senior Scientist in the Imperial Cancer Research Fund Developmental Biology Unit, University of Oxford. Martin Raff received his M.D. from McGill University and is currently a Professor in the MRC Laboratory for Molecular Cell Biology and the Biology Department, University College, London. Keith Roberts received his Ph.D. from the University of Cambridge and is currently Head of the Department of Cell Biology, the John Innes Institute, Norwich. James D. Watson received his Ph.D. from Indiana University and is currently Director of the Cold Spring Harbor Laboratory. He is the author of *Molecular Biology of the Gene* and, with Francis Crick and Maurice Wilkins, won the Nobel Prize in Medicine and Physiology in 1962.

© 1983, 1989, 1994 by Bruce Alberts, Dennis Bray, Julian Lewis, Martin Raff, Keith Roberts, and James D. Watson.

All rights reserved. No part of this book covered by the copyright hereon may be reproduced or used in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or information storage and retrieval systems—without permission of the publisher.

## Library of Congress Cataloging-in-Publication Data

Molecular biology of the cell / Bruce Alberts . . . [et al.] . 3rd ed.  
 p. cm.

Includes bibliographical references and index.

ISBN 0-3153-0619-1 (hard cover) ISBN 0-3153-1620-3 (pbk.)

1. Cytology. 2. Molecular biology. I. Alberts, Bruce.

II. DNLNLF 1. Cells. 2. Molecular Biology. QH1531.2 M713 1994

QH1531.2 M713 1994

574.35—dc20

DNLNLF101

for Library of Congress

93-15907

CIP

Published by Garland Publishing, Inc.  
 711 Fifth Avenue, New York, NY 10022

Printed in the United States of America

15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

Front cover: The photograph shows a cat nerve cell in culture. It is labeled with a fluorescent antibody that stains its cell body and dendritic processes (yellow). Nerve terminals (green) from other neurons (not visible), which have made synapses on the cell, are labeled with a different antibody. (Courtesy of Olaf Mundigl and Pietro de Camilli.)

Dedication page: Gaven Borden, late president of Garland Publishing, weathered in during his mid-1980s climb near Mount McKinley with AlBot (author Bruce Alberts) and famous mountaineer guide Mugs Stump (1940-1992).

Back cover: The authors, in alphabetical order, crossing Abbey Road in London on their way to lunch. Much of this third edition was written in a house just around the corner. (Photograph by Richard Oliver.)

extracts. If these minor cell proteins differ among cells to the same extent as the more abundant proteins, as is commonly assumed, only a small number of protein differences (perhaps several hundred) suffice to create very large differences in cell morphology and behavior.

### A Cell Can Change the Expression of Its Genes in Response to External Signals<sup>3</sup>

Most of the specialized cells in a multicellular organism are capable of altering their patterns of gene expression in response to extracellular cues. If a liver cell is exposed to a glucocorticoid hormone, for example, the production of several specific proteins is dramatically increased. Glucocorticoids are released during periods of starvation or intense exercise and signal the liver to increase the production of glucose from amino acids and other small molecules; the set of proteins whose production is induced includes enzymes such as tyrosine aminotransferase, which helps to convert tyrosine to glucose. When the hormone is no longer present, the production of these proteins drops to its normal level.

Other cell types respond to glucocorticoids in different ways. In fat cells, for example, the production of tyrosine aminotransferase is reduced, while some other cell types do not respond to glucocorticoids at all. These examples illustrate a general feature of cell specialization—different cell types often respond in different ways to the same extracellular signal. Underlying this specialization are features that do not change, which give each cell type its permanently distinctive character. These features reflect the persistent expression of different sets of genes.

### Gene Expression Can Be Regulated at Many of the Steps in the Pathway from DNA to RNA to Protein<sup>4</sup>

If differences between the various cell types of an organism depend on the particular genes that the cells express, at what level is the control of gene expression exercised? There are many steps in the pathway leading from DNA to protein, and all of them can in principle be regulated. Thus a cell can control the proteins it makes by (1) controlling when and how often a given gene is transcribed (**transcriptional control**), (2) controlling how the primary RNA transcript is spliced or otherwise processed (**RNA processing control**), (3) selecting which completed mRNAs in the cell nucleus are exported to the cytoplasm (**RNA transport control**), (4) selecting which mRNAs in the cytoplasm are translated by ribosomes (**translational control**), (5) selectively destabilizing certain mRNA molecules in the cytoplasm (**mRNA degradation control**), or (6) selectively activating, inactivating, or compartmentalizing specific protein molecules after they have been made (**protein activity control**) (Figure 9-2).

For most genes transcriptional controls are paramount. This makes sense because, of all the possible control points illustrated in Figure 9-2, only transcriptional control ensures that no superfluous intermediates are synthesized. In the

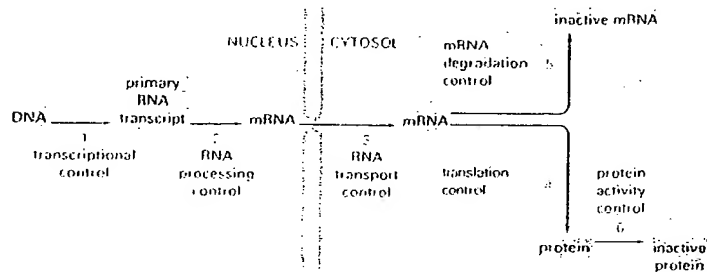


Figure 9-2 Six steps at which eucaryote gene expression can be controlled. Only controls that operate at steps 1 through 5 are discussed in this chapter. The regulation of protein activity (step 6) is discussed in Chapter 5; this includes reversible activation or inactivation by protein phosphorylation as well as irreversible inactivation by proteolytic degradation.

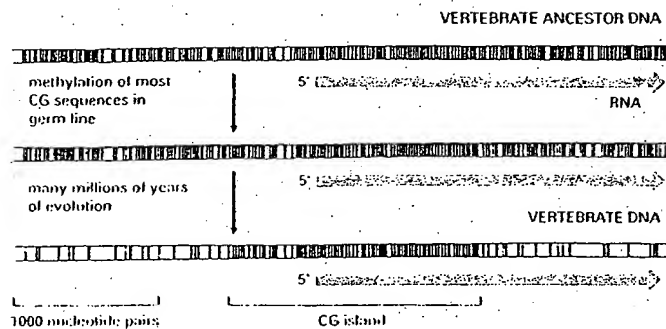


Figure 9-71 A mechanism to explain both the marked deficiency of CG sequences and the presence of CG islands in vertebrate genomes. A black line marks the location of an unmethylated CG dinucleotide in the DNA sequence, while a red line marks the location of a methylated CG dinucleotide.

## Summary

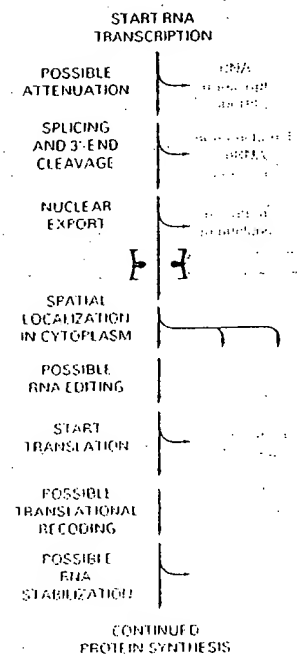
The many types of cells in animals and plants are created largely through mechanisms that cause different genes to be transcribed in different cells. Since many specialized animal cells can maintain their unique character when grown in culture, the gene regulatory mechanisms involved in creating them must be stable once established and heritable when the cell divides, endowing the cell with a memory of its developmental history. Prokaryotes and yeasts provide unusually accessible model systems in which to study gene regulatory mechanisms, some of which may be relevant to the creation of specialized cell types in higher eucaryotes. One such mechanism involves a competitive interaction between two (or more) gene regulatory proteins, each of which inhibits the synthesis of the other; this can create a flip-flop switch that switches a cell between two alternative patterns of gene expression. Direct or indirect positive feedback loops, which enable gene regulatory proteins to perpetuate their own synthesis, provide a general mechanism for cell memory.

In eucaryotes gene transcription is generally controlled by combinations of gene regulatory proteins. It is thought that each type of cell in a higher eucaryotic organism contains a specific combination of gene regulatory proteins that ensures the expression of only those genes appropriate to that type of cell. A given gene regulatory protein may be expressed in a variety of circumstances and typically is involved in the regulation of many genes.

In addition to diffusible gene regulatory proteins, inherited states of chromatin condensation are also utilized by eucaryotic cells to regulate gene expression. In vertebrates DNA methylation also plays a part, mainly as a device to reinforce decisions about gene expression that are made initially by other mechanisms.

Although controls on the initiation of gene transcription are the predominant form of regulation for most genes, other controls can act later in the pathway from RNA to protein to modulate the amount of gene product that is made. Although these posttranscriptional controls, which operate after RNA polymerase has bound to the gene's promoter and begun RNA synthesis, are less common than transcriptional control, for many genes they are crucial. It seems that every step in gene expression that could be controlled in principle is likely to be regulated under some circumstances for some genes.

We consider the varieties of posttranscriptional regulation in temporal order, according to the sequence of events that might be experienced by an RNA molecule after its transcription has begun (Figure 9-72).



Possible post-transcriptional controls on gene expression. Only a few of these controls are likely to be used for any one gene.

MOLECULAR BIOLOGY OF  
**THE CELL**

fourth edition

Bruce Alberts

Alexander Johnson

Julian Lewis

Martin Raff

Keith Roberts

Peter Walter

 **Garland Science**  
Taylor & Francis Group

---

**Garland**

Vice President: Denise Schanck  
 Managing Editor: Sarah Gibbs  
 Senior Editorial Assistant: Kirsten Jenner  
 Managing Production Editor: Emma Hunt  
 Proofreader and Layout: Emma Hunt  
 Production Assistant: Angela Bennett  
 Text Editors: Marjorie Singer Anderson and Betsy Dilernia  
 Copy Editor: Bruce Goatly  
 Word Processors: Fran Dependahl, Misty Landers and Carol Winter  
 Designer: Blink Studio, London  
 Illustrator: Nigel Orme  
 Indexer: Janine Ross and Sherry Granum  
 Manufacturing: Nigel Eyre and Marion Morrow

**Cell Biology Interactive**

Artistic and Scientific Direction: Peter Walter  
 Narrated by: Julie Theriot  
 Production, Design, and Development: Mike Morales

Bruce Alberts received his Ph.D. from Harvard University and is President of the National Academy of Sciences and Professor of Biochemistry and Biophysics at the University of California, San Francisco. Alexander Johnson received his Ph.D. from Harvard University and is a Professor of Microbiology and Immunology at the University of California, San Francisco. Julian Lewis received his D.Phil. from the University of Oxford and is a Principal Scientist at the Imperial Cancer Research Fund, London. Martin Raff received his M.D. from McGill University and is at the Medical Research Council Laboratory for Molecular Cell Biology and Cell Biology Unit and in the Biology Department at University College London. Keith Roberts received his Ph.D. from the University of Cambridge and is Associate Research Director at the John Innes Centre, Norwich. Peter Walter received his Ph.D. from The Rockefeller University in New York and is Professor and Chairman of the Department of Biochemistry and Biophysics at the University of California, San Francisco, and an Investigator of the Howard Hughes Medical Institute.

© 2002 by Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter.  
 © 1983, 1989, 1994 by Bruce Alberts, Dennis Bray, Julian Lewis, Martin Raff, Keith Roberts, and James D. Watson.

All rights reserved. No part of this book covered by the copyright hereon may be reproduced or used in any form in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or information storage and retrieval systems—without permission of the publisher.

**Library of Congress Cataloging-in-Publication Data**

Molecular biology of the cell / Bruce Alberts ... [et al.]. -- 4th ed.  
 p. cm  
 Includes bibliographical references and index.  
 ISBN 0-8153-3218-1 (hardbound) -- ISBN 0-8153-4072-9 (pbk.)  
 1. Cytology. 2. Molecular biology. I. Alberts, Bruce.  
 (DNLM: 1. Cells. 2. Molecular Biology.)  
 QH138J.2 .M64 2002  
 571.6--dc21

2001051471 CIP

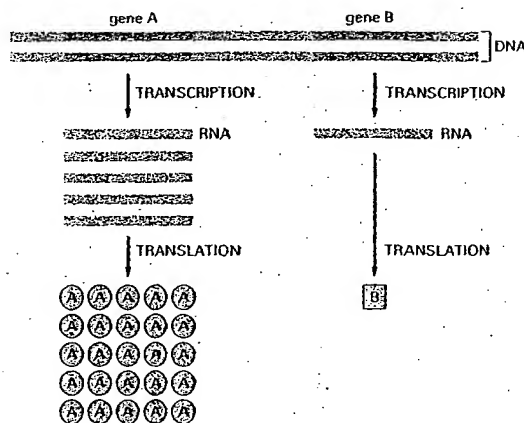
Published by Garland Science, a member of the Taylor & Francis Group,  
 29 West 35th Street, New York, NY 10001-2299

Printed in the United States of America

15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

Front cover Human Genome: Reprinted by permission from *Nature*, International Human Genome Sequencing Consortium, 409:860-921, 2001 © Macmillan Magazines Ltd. Adapted from an image by Francis Collins, NHGRI; Jim Kent, UCSC; Ewan Birney, EBI; and Darryl Leja, NHGRI; showing a portion of Chromosome 1 from the initial sequencing of the human genome.

Back cover In 1967, the British artist Peter Blake created a design classic. Nearly 35 years later Nigel Orme (illustrator), Richard Denyer (photographer), and the authors have together produced an affectionate tribute to Mr Blake's image. With its gallery of icons and influences, its assembly created almost as much complexity, intrigue and mystery as the original, *Drosophila*, *Arabidopsis*, Dolly and the assembled company tempt you to dip inside where, as in the original, "a splendid time is guaranteed for all." (Gunter Blobel, courtesy of The Rockefeller University; Marie Curie, Keystone Press Agency Inc; Darwin bust, by permission of the President and Council of the Royal Society; Rosalind Franklin, courtesy of Cold Spring Harbor Laboratory Archives; Dorothy Hodgkin, © The Nobel Foundation, 1964; James Joyce, etching by Peter Blake; Robert Johnson, photo booth self-portrait early 1930s; © 1926 Delta Haze Corporation all rights reserved, used by permission; Albert L. Lehninger, unidentified photographer; courtesy of The Alan Mason Chesney Medical Archives of The Johns Hopkins Medical Institutions; Linus Pauling, from Ava Helen and Linus Pauling Papers, Special Collections, Oregon State University; Nicholas Proust, courtesy of ArtToday.com; Barbara McClintock, © David Mickles, 1983; Andrei Solhane, courtesy of Elena Bonner, Frederic Singer, © The Nobel Foundation, 1958.)



**Figure 6-3** Genes can be expressed with different efficiencies. Gene A is transcribed and translated much more efficiently than gene B. This allows the amount of protein A in the cell to be much greater than that of protein B.

### FROM DNA TO RNA

Transcription and translation are the means by which cells read out, or express, the genetic instructions in their genes. Because many identical RNA copies can be made from the same gene, and each RNA molecule can direct the synthesis of many identical protein molecules, cells can synthesize a large amount of protein rapidly when necessary. But each gene can also be transcribed and translated with a different efficiency, allowing the cell to make vast quantities of some proteins and tiny quantities of others (Figure 6-3). Moreover, as we see in the next chapter, a cell can change (or regulate) the expression of each of its genes according to the needs of the moment—most obviously by controlling the production of its RNA.

### Portions of DNA Sequence Are Transcribed into RNA

The first step a cell takes in reading out a needed part of its genetic instructions is to copy a particular portion of its DNA nucleotide sequence—a gene—into an RNA nucleotide sequence. The information in RNA, although copied into another chemical form, is still written in essentially the same language as it is in DNA—the language of a nucleotide sequence. Hence the name *transcription*.

Like DNA, RNA is a linear polymer made of four different types of nucleotide subunits linked together by phosphodiester bonds (Figure 6-4). It differs from DNA chemically in two respects: (1) the nucleotides in RNA are *ribonucleotides*—that is, they contain the sugar ribose (hence the name *ribonucleic acid*) rather than deoxyribose; (2) although, like DNA, RNA contains the bases adenine (A), guanine (G), and cytosine (C), it contains the base uracil (U) instead of the thymine (T) in DNA. Since U, like T, can base-pair by hydrogen-bonding with A (Figure 6-5), the complementary base-pairing properties described for DNA in Chapters 4 and 5 apply also to RNA (in RNA, G pairs with C, and A pairs with U). It is not uncommon, however, to find other types of base pairs in RNA: for example, G pairing with U occasionally.

Despite these small chemical differences, DNA and RNA differ quite dramatically in overall structure. Whereas DNA always occurs in cells as a double-stranded helix, RNA is single-stranded. RNA chains therefore fold up into a variety of shapes, just as a polypeptide chain folds up to form the final shape of a protein (Figure 6-6). As we see later in this chapter, the ability to fold into complex three-dimensional shapes allows some RNA molecules to have structural and catalytic functions.

### Transcription Produces RNA Complementary to One Strand of DNA

All of the RNA in a cell is made by DNA transcription, a process that has certain similarities to the process of DNA replication discussed in Chapter 5.

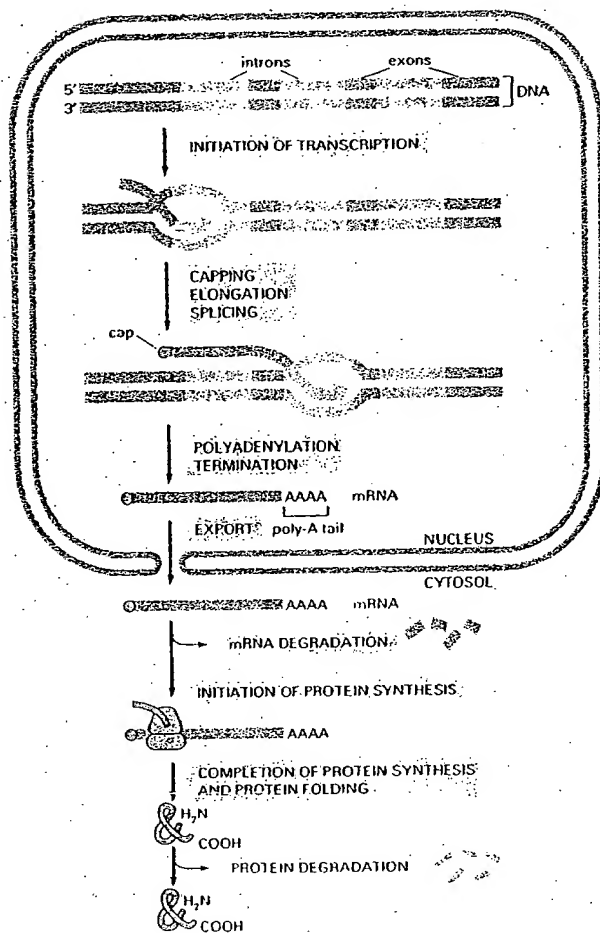


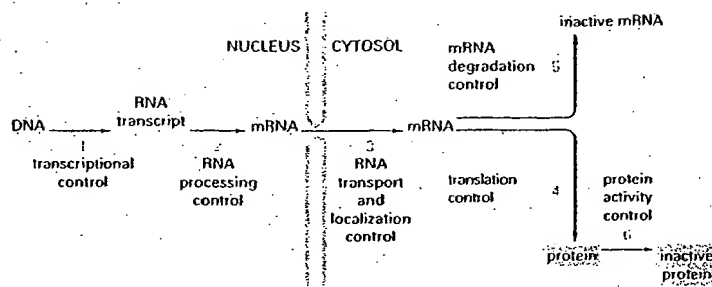
Figure 6-90 The production of a protein by a eucaryotic cell. The final level of each protein in a eucaryotic cell depends upon the efficiency of each step depicted.

ure 6-90) could be regulated by the cell for each individual protein. However, as we shall see in Chapter 7, the initiation of transcription is the most common point for a cell to regulate the expression of each of its genes. This makes sense, inasmuch as the most efficient way to keep a gene from being expressed is to block the very first step—the transcription of its DNA sequence into an RNA molecule.

### Summary

The translation of the nucleotide sequence of an mRNA molecule into protein takes place in the cytoplasm on a large ribonucleoprotein assembly called a ribosome. The amino acids used for protein synthesis are first attached to a family of tRNA molecules, each of which recognizes, by complementary base-pair interactions, particular sets of three nucleotides in the mRNA (codons). The sequence of nucleotides in the mRNA is then read from one end to the other in sets of three according to the genetic code.

To initiate translation, a small ribosomal subunit binds to the mRNA molecule at a start codon (AUG) that is recognized by a unique initiator tRNA molecule. A large ribosomal subunit binds to complete the ribosome and begin the elongation phase of protein synthesis. During this phase, aminoacyl tRNAs—each bearing a specific amino acid—bind sequentially to the appropriate codon in mRNA by forming complementary base pairs with the tRNA anticodon. Each amino acid is added to the C-terminal end of the growing polypeptide by means of a cycle of three sequential



**Figure 7-5** Six steps at which eucaryotic gene expression can be controlled. Controls that operate at steps 1 through 5 are discussed in this chapter. Step 6, the regulation of protein activity, includes reversible activation or inactivation by protein phosphorylation (discussed in Chapter 3) as well as irreversible inactivation by proteolytic degradation (discussed in Chapter 6).

### Gene Expression Can Be Regulated at Many of the Steps in the Pathway from DNA to RNA to Protein

If differences among the various cell types of an organism depend on the particular genes that the cells express, at what level is the control of gene expression exercised? As we saw in the last chapter, there are many steps in the pathway leading from DNA to protein, and all of them can in principle be regulated. Thus a cell can control the proteins it makes by (1) controlling when and how often a given gene is transcribed (**transcriptional control**), (2) controlling how the RNA transcript is spliced or otherwise processed (**RNA processing control**), (3) selecting which completed mRNAs in the cell nucleus are exported to the cytosol and determining where in the cytosol they are localized (**RNA transport and localization control**), (4) selecting which mRNAs in the cytoplasm are translated by ribosomes (**translational control**), (5) selectively destabilizing certain mRNA molecules in the cytoplasm (**mRNA degradation control**), or (6) selectively activating, inactivating, degrading, or compartmentalizing specific protein molecules after they have been made (**protein activity control**) (Figure 7-5).

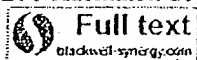
For most genes transcriptional controls are paramount. This makes sense because, of all the possible control points illustrated in Figure 7-5, only transcriptional control ensures that the cell will not synthesize superfluous intermediates. In the following sections we discuss the DNA and protein components that perform this function by regulating the initiation of gene transcription. We shall return at the end of the chapter to the additional ways of regulating gene expression.

#### Summary

*The genome of a cell contains in its DNA sequence the information to make many thousands of different protein and RNA molecules. A cell typically expresses only a fraction of its genes, and the different types of cells in multicellular organisms arise because different sets of genes are expressed. Moreover, cells can change the pattern of genes they express in response to changes in their environment, such as signals from other cells. Although all of the steps involved in expressing a gene can in principle be regulated, for most genes the initiation of RNA transcription is the most important point of control.*

How does a cell determine which of its thousands of genes to transcribe? As mentioned briefly in Chapters 4 and 6, the transcription of each gene is controlled by a regulatory region of DNA relatively near the site where transcription begins. Some regulatory regions are simple and act as switches that are thrown by a single signal. Many others are complex and act as tiny microprocessors, responding to a variety of signals that they interpret and integrate to switch the neighboring gene on or off. Whether complex or simple, these switching devices





### Selective apoptosis of natural killer-cell tumours by l-asparaginase.

Ando M, Sugimoto K, Kitoh T, Sasaki M, Mukai K, Ando J, Egashira M, Schuster SM, Oshimi K.

Department of Haematology, Juntendo University School of Medicine, Tokyo, Japan.

We examined the effectiveness of various anti-tumour agents to natural killer (NK)-cell tumour cell lines and samples, which are generally resistant to chemotherapy, using flow cytometric terminal deoxynucleotidyl transferase-mediated dUTP-biotin nick end-labelling (TUNEL) assay. Although NK-YS and NK-92 were highly resistant to various anti-tumour agents, l-asparaginase induced apoptosis in these two NK-cell lines. NK-cell leukaemia/lymphoma and acute lymphoblastic leukaemia (ALL) samples were selectively sensitive to l-asparaginase and to doxorubicin (DXR) respectively. Samples of chronic NK lymphocytosis, an NK-cell disorder with an indolent clinical course, were resistant to both drugs. Our study clearly separated two major categories of NK-cell disorders and ALL according to the sensitivity to DXR and l-asparaginase. We examined asparagine synthetase levels by real-time quantitative polymerase chain reaction (RQ-PCR) and immunostaining in these samples. At least in nasal-type NK-cell lymphoma, there was a good correlation among asparagine synthetase expression, in vitro sensitivity and clinical response to l-asparaginase. In aggressive NK-cell leukaemia, although asparagine synthetase expression was high at both mRNA and protein levels, l-asparaginase induced considerable apoptosis. Furthermore, samples of each disease entity occupied a distinct area in two-dimensional plotting with asparagine synthetase mRNA level (RQ-PCR) and in vitro l-asparaginase sensitivity (TUNEL assay). We confirmed rather specific anti-tumour activity of l-asparaginase against NK-cell tumours in vitro, which provides an experimental background to the clinical use of l-asparaginase for NK-cell tumours.

PMID: 16156856 [PubMed - in process]

**Human thyroid carcinoma cell lines and normal thyrocytes: expression and regulation of matrix metalloproteinase-1 and tissue matrix metalloproteinase inhibitor-1 messenger-RNA and protein.**

Aust G, Hofmann A, Laue S, Rost A, Kohler T, Scherbaum WA.

Institut of Anatomy, University of Leipzig, Germany.

Matrix metalloproteinase-1 (MMP-1) and tissue matrix metalloproteinase inhibitor 1 (TIMP-1) play an important role in remodeling the extracellular matrix in normal and pathological processes. The effect of phorbol-myristate acetate (PMA), interleukin-1 (IL-1), and tumor necrosis factor-alpha (TNF-alpha) on MMP-1 and TIMP-1 expression was studied on highly purified thyrocytes and undifferentiated 8505 C, C 643, HTh 74, SW 1736 thyroid carcinoma cells compared with thyroid-derived fibroblasts. Messenger RNA (mRNA) levels were monitored by competitive semiquantitative reverse transcriptase polymerase chain reaction (RT-PCR) after 24 hours. Culture supernatants were assayed for free and/or complexed MMP-1 and TIMP-1 after 48 hours using enzyme-linked immunosorbent assay (ELISA) systems (detection limit: <2 ng/mL). MMP-1 and TIMP-1 mRNA were present in all cell types, although thyrocytes showed MMP-1 mRNA levels near the detection limit. 8505 C expressed MMP-1 mRNA levels of up to 10(6) times those of the other cells analyzed. PMA and IL-1 increased MMP-1 mRNA in most cell types. TIMP-1 mRNA increased after treatment with PMA in all cells except 8505 C, whereas only slight effects were shown after IL-1 stimulation. MMP-1 protein was undetectable in normal thyrocyte cultures, but was secreted spontaneously by all cell lines ([ng/mL]; C 643: 15+/-7; HTh 74: 81+/-1; SW 1736: 13+/-2; 8505 C: 2097+/-320). There was a strong correlation between levels of MMP-1 mRNA and protein ( $r = 0.99$ ,  $p < .0001$ ). PMA and IL-1 increased MMP-1 secretion in all cell types after 48 hours. Fibroblasts ([ng/mL] 517+/-55) and the cell lines (C 643: 142+/-48; HTh 74: 115+/-13; SW 1736: 202+/-14; 8505C: 120+/-19) secreted TIMP-1 in unstimulated cultures, whereas only a trace amount was detected in thyrocyte cultures, even after PMA treatment. IL-1 upregulated TIMP-1 secretion after 48 hours in SW 1736, HTh 74, and C 643 cells. Our data suggest that in contrast to normal thyrocytes, dedifferentiated thyroid carcinoma cell lines are potential producers of MMP-1 as well as TIMP-1. High MMP-1 or MMP-1/TIMP-1 expression may play a role in tissue invasion of undifferentiated thyroid cancer cells.

PMID: 9349574 [PubMed - indexed for MEDLINE]

# Thyroid

VOLUME 7

NUMBER 5

OCTOBER 1997

## CLINICAL RESEARCH

- Levothyroxine Suppressive Therapy is Partially Effective in Treating Patients with Benign, Solid Thyroid Nodules and Multinodular Goiters.  
NICOLAU LIMA, MEYER KNOBEL, HUMBERTO CAVALIERE,  
CLAUDIA SZTEJNSZNAJD, EDUARDO TOMIMORI, and  
GERALDO MEDEIROS-NETO 691
- Is Percutaneous Ethanol Injection a Useful Alternative for the Treatment of the Cold Benign Thyroid Nodule? Five Years' Experience.  
NADIA CARACCIO, ORLANDO GOLETTI, PIERO VINCENZO LIPPOLIS,  
ARTURO CASOLARO, ENRICO CAVINA, PAOLO MICCOLI, and FABIO MONZANI 699
- Value of Combined Technetium-99m Hydroxy Methylene Diphosphonate and Thallium-201 Imaging in Detecting Bone Metastases from Thyroid Carcinoma.  
MD. SAYEEDUL ALAM, RYO TAKEUCHI, KANJI KASAGI, TAKASHI MISAKI,  
SHINICHI MIYAMOTO, YASUHIRO IIDA, AKINARI HIDAKA, and JUNJI KONISHI 705
- Human Thyroid Carcinoma Cell Lines and Normal Thyrocytes: Expression and Regulation of Matrix Metalloproteinase-1 and Tissue Matrix Metalloproteinase Inhibitor-1 Messenger-RNA and Protein.  
G. AUST, A. HOFMANN, S. LAUE, A. ROST, T. KÖHLER, and W.A. SCHERBAUM 713
- MUC1 Mucin Gene, Transcripts, and Protein in Adenomas and Papillary Carcinomas of the Thyroid.  
IVAN BIÈCHE, EMMANUEL RUFFET, ALAIN ZWEIBAUM, FRANÇOISE VILDÉ,  
ROSETTE LIDEREAU, and BRIGITTE FRANC 725
- Incidence and Clinical Characteristics of Thyroid Carcinoma After Iodine Prophylaxis in an Endemic Goiter Country.  
C. BACHER-STIER, G. RICCABONA, M. TÖTSCH, G. KEMMLER, W. OBERAIGNER,  
and R. MONCAYO 733
- Opposite Changes in Serum Soluble CD8 in Patients at the Active Stages of Graves' and Hashimoto's Diseases.  
MIKIO WATANABE, NOBUYUKI AMINO, KAZUNORI HOCHITO,  
KIYOSHI WATANABE, KANJI KUMA, and YOSHINORI IWATANI 743
- Urinary Iodine Excretion During Normal Pregnancy in Healthy Women Living in the Southwest of France: Correlation with Maternal Thyroid Parameters.  
PHILIPPE CARON, MADELEINE HOFF, SAMUEL BAZZI, ALAIN DUFOR,  
GÉRARD FAURE, IMAD GHANDOUR, PATRICK LAUZU, YVAN LUCAS,  
DOMINIQUE MARAVAL, FRÉDÉRIC MIGNOT, PASCAL RÉSSIGÉAC,  
FRANÇOISE VERTONGEN, and VÉRONIQUE GRANGÉ 749

(continued)

## Human Thyroid Carcinoma Cell Lines and Normal Thyrocytes: Expression and Regulation of Matrix Metalloproteinase-1 and Tissue Matrix Metalloproteinase Inhibitor-1 Messenger-RNA and Protein

G. AUST,<sup>1</sup> A. HOFMANN,<sup>2</sup> S. LAUE,<sup>2</sup> A. ROST,<sup>3</sup> T. KÖHLER,<sup>3</sup> and W.A. SCHERBAUM<sup>4</sup>

### ABSTRACT

Matrix metalloproteinase-1 (MMP-1) and tissue matrix metalloproteinase inhibitor 1 (TIMP-1) play an important role in remodeling the extracellular matrix in normal and pathological processes. The effect of phorbol-myristate acetate (PMA), interleukin-1 (IL-1), and tumor necrosis factor- $\alpha$  (TNF- $\alpha$ ) on MMP-1 and TIMP-1 expression was studied on highly purified thyrocytes and undifferentiated 8505 C, C 643, HTh 74, SW 1736 thyroid carcinoma cells compared with thyroid-derived fibroblasts. Messenger RNA (mRNA) levels were monitored by competitive semiquantitative reverse transcriptase polymerase chain reaction (RT-PCR) after 24 hours. Culture supernatants were assayed for free and/or complexed MMP-1 and TIMP-1 after 48 hours using enzyme-linked immunosorbent assay (ELISA) systems (detection limit:  $<2$  ng/mL). MMP-1 and TIMP-1 mRNA were present in all cell types, although thyrocytes showed MMP-1 mRNA levels near the detection limit. 8505 C expressed MMP-1 mRNA levels of up to  $10^6$  times those of the other cells analyzed. PMA and IL-1 increased MMP-1 mRNA in most cell types. TIMP-1 mRNA increased after treatment with PMA in all cells except 8505 C, whereas only slight effects were shown after IL-1 stimulation. MMP-1 protein was undetectable in normal thyrocyte cultures, but was secreted spontaneously by all cell lines ([ng/mL]: C 643:  $15 \pm 7$ ; HTh 74:  $81 \pm 1$ ; SW 1736:  $13 \pm 2$ ; 8505C:  $2097 \pm 320$ ). There was a strong correlation between levels of MMP-1 mRNA and protein ( $r = 0.99$ ,  $p < .0001$ ). PMA and IL-1 increased MMP-1 secretion in all cell types after 48 hours. Fibroblasts ([ng/mL]  $517 \pm 55$ ) and the cell lines (C 643:  $142 \pm 48$ ; HTh 74:  $115 \pm 13$ ; SW 1736:  $202 \pm 14$ ; 8505 C:  $120 \pm 19$ ) secreted TIMP-1 in unstimulated cultures, whereas only a trace amount was detected in thyrocyte cultures, even after PMA treatment. IL-1 upregulated TIMP-1 secretion after 48 hours in SW 1736, HTh 74, and C 643 cells. Our data suggest that in contrast to normal thyrocytes, dedifferentiated thyroid carcinoma cell lines are potential producers of MMP-1 as well as TIMP-1. High MMP-1 or MMP-1/TIMP-1 expression may play a role in tissue invasion of undifferentiated thyroid cancer cells.

### INTRODUCTION

MATRIX METALLOPROTEINASES, (MMPs) constitute a family of structurally related proteolytic enzymes responsible for the proteolytic degradation of extracellular matrix (ECM) components. They are important participants in normal tissue remodeling and contribute to the phenotype of several pathological conditions that are associated with progressive ECM degradation. MMPs are highly regulated at different levels (1). At the transcriptional level, MMP expression can be directly induced or

suppressed on external stimulation, ie, with cytokines, phorbol 12-myristate 13-acetate (PMA), lipopolysaccharide (LPS), or retinoic acid (2,3). After secretion at post-transcriptional level, latent MMP proenzymes are regulated by proteolytic activation and interaction with tissue inhibitors of matrix metalloproteinase (TIMPs), their specific inhibitors. Any imbalance between the proteolytic MMPs activities and the TIMPs that could be influenced and caused by cytokines could potentially lead to pathological conditions (4).

MMP-1, although known as an interstitial collagenase,

<sup>1</sup>Institut of Anatomy, <sup>2</sup>Department of Internal Medicine III, and <sup>3</sup>Institute of Clinical Chemistry and Pathobiochemistry, University of Leipzig, Germany. <sup>4</sup>Department of Endocrinology, University of Düsseldorf.

is the only enzyme active at neutral pH that can degrade extracellular fibers comprised of collagen types I, II, and III. With this initial step, MMP-1 provides the cleavage products to other collagenase types (5). The major specific inhibitor of MMP-1 is TIMP-1, a 28.5-kd glycoprotein, which forms 1:1 stoichiometric complexes with the protease (6). Cytokines and growth factors have been shown to regulate the expression of both MMP-1 and TIMP-1 (1,7,8).

Although the participation of MMP-1 as the initial collagenase in tissue breakdown during tumor development is well documented (9-11), only one study has described the expression (12) but no study has as yet investigated the regulation of this enzyme in different thyroid tumors. Few studies have been published investigating the role of other MMPs in normal and pathological thyroid tissue by *in situ* hybridization and immunohistochemistry (13-16). Furthermore, tissue remodeling includes both the action of MMPs and their inhibitors; thus, these enzymes could be involved in autoimmune and other nonautoimmune thyroid diseases during morphological changes (17,18). It is still unknown whether or not thyrocytes are able to express MMPs and TIMPs. Although type IV collagenases (MMP-2 and MMP-9) were detected in various human epithelial cells of different tissue origin (19,20), only one study described the secretion of MMP-1 by epithelial cells (21).

Highly purified normal thyrocytes and four thyroid carcinoma cell lines were included in this study to investigate the involvement of these cells in MMP-1 and TIMP-1 production during thyroid tissue remodeling processes and in malignant thyroid neoplasms. MMP-1 and TIMP-1 expression were studied at both the mRNA and protein level by semiquantitative RT-PCR and ELISA measurement, respectively.

In unstimulated carcinoma cell lines both MMP-1 and TIMP-1 mRNA were expressed, partly at a high level, followed by the spontaneous secretion of the proteins. The various conditions for the stimulation of the different cell lines by cytokines and PMA were defined. In contrast to the cell lines, normal thyrocytes did not secrete MMP-1 and only trace amount of TIMP-1, even after stimulation with PMA.

## MATERIALS AND METHODS

### Preparation of tissues, thyroid-derived cells, and cell lines

Thyrocytes were prepared from surgical thyroid specimens from 3 patients (1 Graves' disease, 2 nontoxic goiter; mean age  $54.3 \pm 5.0$  years). Fibroblasts were separated from thyroid tissue of 5 other patients (3 Graves' disease, 2 nontoxic goiter; mean age  $43.6 \pm 6.4$  years). Graves' disease and nontoxic goiter were diagnosed on the strength of clinical, biochemical, and immunologic features as well as thyroid scintiscans.

Thyroid tissue was trimmed of fat and connective tissue immediately after surgery. Thyroid-derived cells were enriched after gradual enzymatic digestion of tissue and cultured over a period of 16 hours as described. Thyrocytes were obtained from the adherent fraction by incubating

the cell monolayer with phosphate buffered saline (PBS) without  $\text{Ca}^{2+}/\text{Mg}^{2+}$  for 45 minutes (22). Residual fibroblasts were removed after subsequent incubation of the cells with the fibroblast-specific mab FibAS01 (22) and goat-anti-mouse IgG-DYNABEADS® M450 (DYNAL, Hamburg, Germany) according to the manufacturer's protocol.

Thyroid-derived fibroblasts were obtained after culturing small pieces of thyroid tissue in Dulbeccos's Modified Eagle's Medium (DMEM) with 10% fetal calf serum (FCS) and harvested in the 5th to 7th passage. The purity of the thyrocytes and fibroblasts was determined by using indirect immunofluorescence technique on a FACS-Scan (Becton Dickinson GmbH, Heidelberg, Germany) as described (22).

The following human anaplastic thyroid carcinoma cell lines were cultured in DMEM with 10% FCS: C 643 (23); SW 1736 (23); and HTh 74 (24). The cell line 8505 C (25) was purchased from the German Collection of Microorganisms and Animal Cell Cultures (DSM ACC219). This cell line was established from a primary thyroid tumor characterized histologically as a undifferentiated carcinoma that was partially composed of poorly differentiated papillary cells (25). This is a feature of a subgroup of anaplastic carcinoma (26). The majority of these coexistent better differentiated carcinoma foci in anaplastic carcinoma were papillary (26).

### In vitro cultures

Using 24-well plates,  $1 \times 10^5$  cells were cultured for 24 hours. The medium was aspirated and replaced with 500  $\mu\text{L}$  OPTI-MEM (GIBCO BRL, Grand Island, NY) without FCS to eliminate possible stimulation of MMP-1 and TIMP-1 production by FCS. The medium contained the desired concentration of human IL-1 $\alpha$  (10 U/mL; Pepro Tech EC Ltd., London, UK), TNF- $\alpha$  (100 U/mL; Pepro Tech EC Ltd.), interferon- $\gamma$  (IFN- $\gamma$ ) (500 U/mL; Pepro Tech EC Ltd.), or 10 ng/mL PMA (SIGMA).

Triplicate cultures of each stimulator were analyzed after 3, 6, and 24 hours at the mRNA and after 24 and 48 hours at the protein level. The supernatants were removed and stored at  $-80^\circ\text{C}$  for further use. First, a collagenolytic assay based on the digestion of type I collagen was performed. This method showed direct evidence of free pro-MMP-1 enzyme in the cell culture supernatants of unstimulated and IL-1 $\alpha$  stimulated 8505 C, HTh 74, and C634 cells (data not shown). However, the method does not allow quantitation of MMP-1 enzyme activity. Thus, the cell culture supernatants were assayed for MMP-1, TIMP-1, and MMP1/TIMP-1 complex by ELISA (Amersham Life Sciences, Braunschweig, Germany). The MMP-1 assay (sensitivity: 1.7 ng/mL) detected only total human MMP-1, ie, free MMP-1 and MMP-1 complexed with inhibitors such as TIMP-1. It did not detect MMP-1 bound by the nonspecific protease inhibitor  $\alpha_2$ -macroglobulin. The MMP-1/TIMP-1 assay (sensitivity: 1.5 ng/mL) detected MMP-1/TIMP-1 complex, ie, activated MMP-1 that has been subsequently complexed with the specific MMP-1 inhibitor TIMP-1. It did not detect free active MMP-1, free TIMP-1, or pro-MMP-1. There was no cross-reactivity with active MMP-1 bound by the nonspecific protease inhibitor  $\alpha_2$ -macroglobulin. The TIMP-1 assay (sensitivity:

1.25 ng/mL) detected total human TIMP-1, ie, free TIMP-1 and that complexed with MMPs. The assay did not fully cross-react with TIMP-1 in complexes with other MMP. It did not cross-react with TIMP-2.

#### RNA extraction and cDNA synthesis

For gene expression studies, 5 mL RNAzol™ B (Biotex Laboratories Inc., Houston, TX) was added to the cell culture wells. The content of three wells of any cell type was pooled and then stored frozen for further mRNA analysis in liquid nitrogen. Total cellular RNA (cDNA) was isolated from the probes according to the manufacturer's protocol. RNA was fractionated on a denaturing 1.0% agarose gel and stained with ethidium bromide to confirm that spectrophotometric measurements were accurate and that the RNA had not been degraded. Five micrograms total RNA was reverse-transcribed to cDNA using the First-strand cDNA synthesis kit of Pharmacia (Uppsala, Sweden) in a total reaction volume of 15  $\mu$ L.

#### mRNA analysis by competitive RT-PCR

To correct for variations across different cDNA preparations, all samples were first adjusted to contain equal input glyceraldehyde-3-phosphate dehydrogenase (GAPDH) cDNA concentrations. Semi-quantitative GAPDH RT-PCR was used with a heterologous synthetic competitor fragment. The generation of the specific PCR products from the competitor and the cDNA with the GAPDH primers were published earlier (22,27).

We then estimated the MMP-1 and TIMP-1 cDNA in these adjusted samples. The primers were selected using the DNAsis computer program (Hitachi Software Engineering Co, Yokohama, Japan). The primer pairs span one or more introns to allow unambiguous discrimination between cDNA and unwelcome contaminating genomic DNA. In quantitating MMP-1 and TIMP-1 cDNA, a rapid one-step method was introduced to synthesize an internal homologous competitor (plan diagram of procedure: Fig. 1, exemplary for MMP-1 [28]). A hybrid primer was synthesized (MMP-1hy) that consisted of two segments (seg<sub>1</sub>, seg<sub>2</sub>). It

had a length of 40 nucleotides, in which 20 nucleotides (seg<sub>1</sub>) at the 3' end corresponded to the opposite strand of the target sequence a predetermined distance from primer MMP-1f, and 20 nucleotides at the 5' end (seg<sub>2</sub> = MMP-1r) that corresponded to the target sequence upstream from the segment seg<sub>1</sub>. Amplification with the primers MMP-1f and MMP-1hy from the cDNA resulted in a 478-base pair (bp) (polymerase chain reaction (PCR) product. It was freed from excess primers and deoxynucleoside-triphosphates (dNTPs) using the Qiaquick Gel Extraction Kit (Qiagen GmbH, Hilden, Germany) and quantified. A known number of copies of the competitor was introduced in the GAPDH-adjusted samples and amplified with the primers MMP-1f and MMP-1r. With this approach, two products were generated, one derived from the cDNA (560 bp) and another, 82 bp smaller in size derived from the internal competitor (Fig. 1). PCR products were resolved by gel electrophoresis (1.5% agarose gel). The relative amounts of sample cDNA and competitor were quantified by measuring the intensity of ethidium fluorescence with a CCD image sensor and analyzing the data with the EASY program (Herolab, Wiesloch, Germany). The initial amounts of sample cDNA and competitor were assumed to be equal in those reactions where the ratio of the two products was judged to be equal. This was expressed in arbitrary units (AU) (22,29). One AU was defined as the lowest concentration of competitor yielding a detectable amplification product when added to PCR alone. For example, if equivalence between sample cDNA and competitor was reached using a 100-fold concentrated competitor the relative sample cDNA concentration was 100 AU. Thyrocytes and the cell lines were analyzed for the expression of thyroid-specific and cytokine receptor mRNAs in a simple RT-PCR. The sequences of the TPO and cytokine receptor primer pairs have been published by Watson et al. (30) and Tada et al. (31) and gave the following product sizes: TPO: 506 bp; IL-1R type I (p80): 300 bp; IL-1R type II (p68): 392 bp; TNF- $\alpha$ R (p75): 324 bp; TNF- $\alpha$ R (p55): 587 bp and IFN- $\gamma$ R: 899 bp. The thyroglobulin (Tg) and thyroid stimulating hormone receptor (TSHR) primer pairs were selected according to the published sequences using the DNAsis program (Table 1).

Each 25- $\mu$ L amplification reaction contained 2.5  $\mu$ L 10  $\times$  concentrated PCR buffer (15 mM MgCl<sub>2</sub>, Boehringer Mannheim, Germany), 0.3 U Taq DNA polymerase (Boehringer Mannheim, Germany), 100  $\mu$ M dNTPs (Perkin Elmer, Weiterstadt, Germany), 0.1  $\mu$ M of each primer (IMB, Jena, Germany), and 1  $\mu$ L cDNA and competitor in adjusted dilution. Furthermore, restriction mapping (restriction enzymes: Boehringer Mannheim GmbH, Germany) was carried out to confirm the originality of the PCR product (Fig. 1, Table 1).

#### Statistics

Protein levels of thyrocyte or fibroblast cultures from the different patients and of the thyroid carcinoma cell lines obtained from three separate experiments were presented as mean  $\pm$  SEM values. Statistical comparisons between unstimulated and stimulated cell cultures were performed by the alternate (Welch) t-test. The correlation between basal mRNA levels and the unstimulated protein secretion in all cell types was calculated according to the Spearman method.

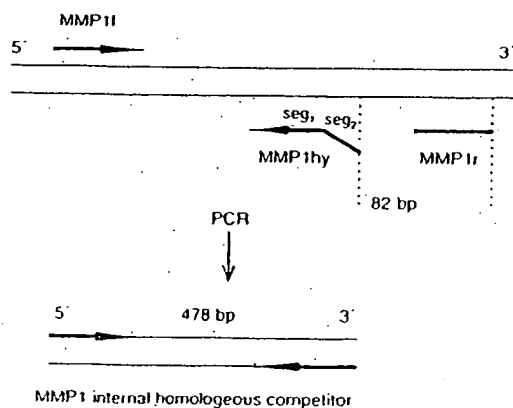


FIG. 1. General scheme for generating homologous competitors used for quantitative PCR.

TABLE 1. PRIMERS, LENGTH OF AMPLIFIED TEMPLATES, RESTRICTION MAPPING AND ASSAY CONDITIONS FOR RT-PCR

		Primer	Length of cDNA (bp)	Length of competitor (bp)	Annealing temperature	Number of cycles
	5'-	3'				
Tg	forward	GCAGATCTTACTGAGTGGCT	416		60	35
	reverse	TGTCAGCACAGTGGCAATAC				
TSH-R	forward	ACTTGCTGCAGCTGGTGCT	354		65	35
exons 1-4	reverse	TGAGGGCATCAGGGTCTATG				
TSH-R	forward	GAAATTCGGAATACCAAGGAACCTA ACT	896		53	35
exons 4-10	reverse	AACTCATCGGACTTGGGGGTACA				
MMP-1	forward	TGGGAGCAAAACACATCTGAC	560	478	64	33
	reverse	ATCATTCTCCCCGAATCGT				
	hybrid	ATCATTCTCCCCGAATCGT				
		CCATATATGGCTTGGATGCC				
TIMP-1	forward	CTTAGGGGATGCCGCTGACA	351	274	64	30
	reverse	GGCAGGCAGGCAAGGTGACG				
	hybrid	GGCAGGCAGGCAAGGTGACG				
		GGATGGATAAACAGGGAAC				

Tg indicates thyroglobulin; TSH-R, thyroid stimulating hormone receptor; MMP-1, matrix metalloproteinase-1; TIMP-1, tissue inhibitor of metalloproteinase-1; bp, base pair.

## RESULTS

### Thyroid specific and cytokine receptor mRNA expression

Isolated thyrocytes as well as 8505 C cells expressed Tg and thyroperoxidase (TPO) mRNA, whereas transcripts of the TSH-R (exons 1-4, 354 bp, exons 4-10, 896 bp) were present only in the thyrocytes. The three anaplastic thyroid carcinoma cell lines SW 1736, C 634, and HTh 74 were completely negative for the Tg, TPO, and TSH-R mRNAs (Fig. 2). All cell lines and thyrocytes expressed IL-1R (type I and type II), TNF- $\alpha$ R (p75 and p55) and IFN- $\gamma$ R mRNA (Fig. 2).

### Basal MMP-1 and TIMP-1 mRNA and protein expression

In most stimulation experiments, mRNA levels did not increase until 24 hours of incubation. The 24-hour mRNA levels are shown in Figures 3 and 4. The 3- and 6-hour levels are demonstrated in those experiments where the mRNA levels reached their peak before 24 hours of stimulation. If not otherwise indicated, the MMP-1 levels were measured using the ELISA system, which recognizes free/complexed MMP-1.

MMP-1 and TIMP-1 mRNA were found during unstimulated culture in all investigated cell types, although the mRNA levels varied over a great range. 8505 C showed a basal MMP-1 mRNA level 20 times as high as those of the HTh 74 cells,  $6 \times 10^4$  times as high as C 634, and  $2 \times 10^6$  times as high as SW 1736 cells. In thyrocytes, MMP-1 mRNA levels were found near the detection limit (Figs. 3 and 4).

Generally, when analyzing the noted cell types, the measured basal MMP-1 or TIMP-1 mRNA levels correlated well with the basal protein expression (MMP-1:  $r = 0.99$ ;  $p < .0001$ ; TIMP-1:  $r = 0.98$ ,  $p < .002$ ). Corresponding to the high MMP-1 mRNA level, 8505 C cells secreted extremely high levels of MMP-1. No MMP-1 or TIMP-1 was

detected in unstimulated thyrocyte cultures at any time-point examined. All other cell types showed a spontaneous MMP-1 and TIMP-1 secretion (Figs. 5 and 6). Thyroid-derived fibroblasts produced basal TIMP-1 levels of up to 4 times higher in the four carcinoma cell lines, which secreted nearly the same amounts of basal TIMP-1 protein. Nevertheless, TIMP-1 secretion of fibroblasts was found at lower levels than expected after TIMP-1 mRNA measurement in 4 of 5 analyzed patients. The results of the fibroblast cultures from patient five showing a higher TIMP-1 expression than those from the 4 other patients (basal 24 hour:  $50 \pm 2$ ; PMA 24 hour:  $90 \pm 6$  ng/mL TIMP-1) was omitted in Figure 6.

Comparing the basal amount of free/complexed and TIMP-1 complexed MMP-1 after 24 hours of stimulation, a significant level of MMP-1 was not complexed with TIMP-1 in 8505 C cultures, whereas in fibroblast cultures most of the MMP-1 activity was inhibited by TIMP-1. The anaplastic carcinoma cell line HTh 74 did not show such a great discrepancy between free/complexed and TIMP-1 complexed MMP-1 level as 8505 C cells (Fig. 7).

### Effects of IL-1 $\alpha$ on MMP-1 and TIMP-1 mRNA and protein expression

Experiments were performed to determine whether human thyroid epithelial cells and thyroid carcinoma cell lines could produce or increase basal MMP-1 and TIMP-1 secretion after exposure to various stimuli. The results from these stimulation experiments are summarized in Figures 3 through 5. Generally, there was a delay in protein secretion level in comparison to the mRNA expression level. At the protein level, the cytokine-mediated stimulating or inhibiting effect is more distinct after 48 hours compared with 24 hours, even when the mRNA level had already decreased after 6 hours.

IL-1 upregulated MMP-1 mRNA in SW 1736 cells up to 100 times and, in thyroid-derived fibroblasts, up to 12 times after 24 hours of incubation (Fig. 3). This increased mRNA level was accompanied by a significantly enhanced

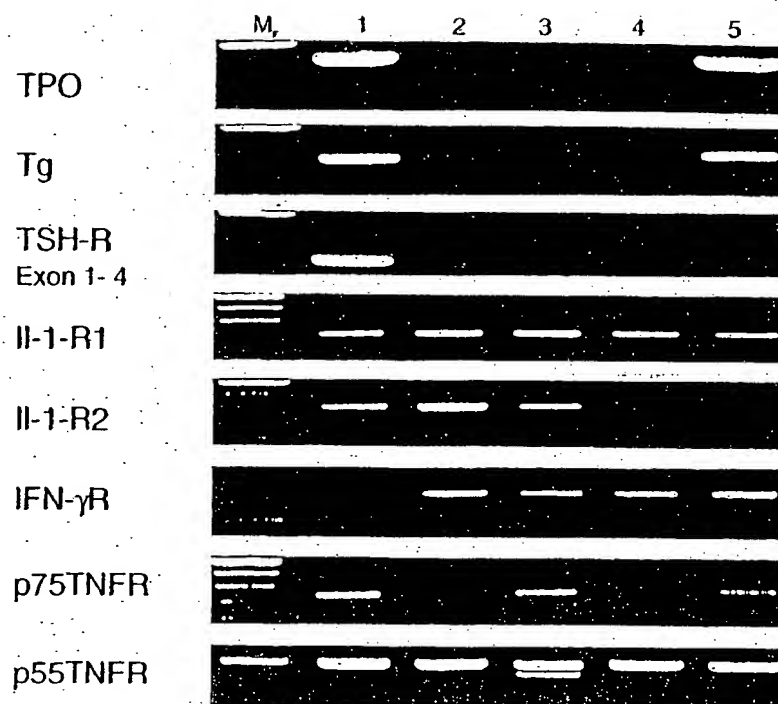


FIG. 2. Amplification of thyroid specific and interleukin-receptor mRNA in thyrocytes (1), SW 1736 (2), C 643 (3), HTh 74 (4) and 8505 C (5) cells using RT-PCR;  $M_r$  = 100-bp ladder (GIBCO).

MMP-1 secretion after 48 hours. Furthermore, IL-1 $\alpha$  increased MMP-1 mRNA expression in thyrocytes up to seven times after 6 hours, but no MMP-1 protein could be detected in thyrocyte cultures. IL-1 had no stimulatory effect on MMP-1 mRNA expression in C 643, HTh 74, and 8505 C cells after 24 hours, although a significant increase of MMP-1 secretion was found in HTh-74 and SW 1736 cells after 48 hours of incubation (Fig. 5). This discrepancy may be explained by a possible increase in MMP-1 mRNA level after 24 hours of stimulation. The same effect could also be observed in the IL-1 stimulated TIMP-1 at the mRNA as well as the protein level: the only slight effect of IL-1 on TIMP-1 mRNA expression in carcinoma cell lines after 24 hours was accompanied by a significant increase of TIMP-1 secretion in 8505 C and HTh 74 cells after 48 hours (Figs. 4 and 6).

#### *Effects of TNF- $\alpha$ on both MMP-1/TIMP-1 mRNA and protein expression*

In contrast to IL-1, TNF- $\alpha$  did not stimulate the MMP-1 and TIMP-1 mRNA and protein levels in all carcinoma cell lines and thyrocytes. Only thyroid-derived fibroblasts responded with a slight upregulation of MMP-1 and TIMP-1 mRNA expression after TNF- $\alpha$  stimulation, which was not accompanied by an increase of MMP-1 and TIMP-1 secretion.

#### *Effects of PMA, and IFN- $\gamma$ on MMP-1 and TIMP-1 mRNA and protein expression*

PMA was included in our study as a positive control because it is known to upregulate or induce both MMP-1 and TIMP-1 secretion in various cell types (1,32). Indeed, PMA was able to induce or enhance MMP-1 mRNA levels in all cell types investigated, although the detected levels varied to a large extent (Fig. 3). This result is in good correlation with the significantly increased MMP-1 protein levels that were already detectable after 24 hours of stimulation (Fig. 5). PMA upregulated TIMP-1 mRNA levels by up to 20 times in C 643, and up to 2 times in SW 1736 and HTh 74 cells, fibroblasts and thyrocytes, but it did not change the TIMP-1 mRNA content in 8505 cells (Fig. 4). At the protein level, we found a significant stimulation of TIMP-1 secretion in C 643 and HTh 74 cells, as well as in thyroid-derived fibroblasts (Fig. 6).

In contrast to PMA, IFN- $\gamma$  was without effect on stimulation or downregulation of MMP-1 and TIMP-1 mRNA or protein in any of the cell types investigated (Figs 5 and 6).

The main inhibitor of MMP-1 is TIMP-1, which forms 1:1 stoichiometric complexes with MMP-1, although some other inhibitors can also bind MMP-1. On the other hand, TIMP-1 can bind other MMP types.



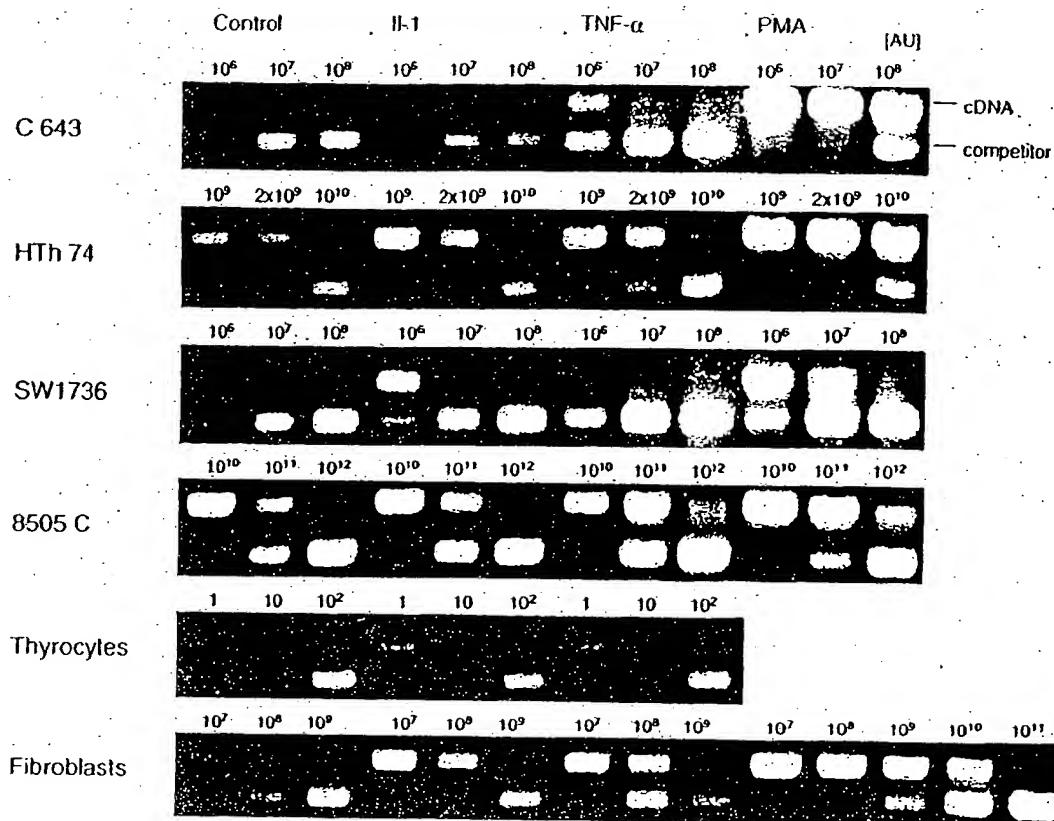


FIG. 3. Representative samples of competitive amplified MMP-1 mRNA of thyrocytes, thyroid-derived fibroblasts and thyroid carcinoma cell lines without stimulation (control) and after stimulation with 10 U/mL IL-1 $\alpha$  and 100 U/mL TNF- $\alpha$  and 10 ng/mL PMA after 24 hours. Serial dilutions of known amounts of the competitor fragment were coamplified with identical aliquots of cDNA. The 560-bp (cDNA) and 478-bp (competitor) PCR products were visualized by agarose gel electrophoresis and ethidiumbromide staining. The relative concentration of the added competitor was given in arbitrary units (AU) in the figure. One AU was defined as the lowest concentration of the competitor yielding a detectable amplification for MMP-1 mRNA. The ratio of competitor to cDNA fragments was determined by measuring the intensity of ethidium fluorescence with a CCD image sensor and analysis of data. Measured cDNA concentration can be expressed in AU.

## DISCUSSION

Our findings demonstrate for the first time that thyroid carcinoma cell lines are able to express MMP-1 and TIMP-1 mRNA and protein at significant levels *in vitro*. The observation of spontaneous release of MMP-1 and TIMP-1 corresponds well with earlier studies covering the secretion of these proteins by several carcinoma cell lines (33,34).

However, in contrast to its clear physiological function in extracellular matrix breakdown, the role of MMP-1 in tumor growth and metastases is still controversial (9-11,35). Recently, Murray et al. (10) demonstrated that MMP-1 is associated with poor prognosis in colorectal cancer, and has a prognostic value independent of the Duke stage. Therefore, MMP-1 could be a target for therapeutic intervention in such tumors. Furthermore, the hypothesis of whether or not cancer cells themselves are able to

produce MMP, or whether cancer cells stimulate the surrounding stromal cells to secrete MMP *in vivo*, is disputed. MMP-1 mRNA and protein were detected by both *in situ* hybridization and immunohistochemistry in stromal as well as tumor cells of head, neck, gastric, colorectal, and mammary carcinomas (9,10,36,37). In contrast, Kameyama (12) demonstrated by *in situ* hybridization that the MMP-1 mRNA was not expressed in the cancer cells but in the surrounding fibrous capsules of strongly differentiated papillary thyroid carcinoma tissue. Highly differentiated follicular carcinomas and follicular adenomas were depleted for MMP-1 transcripts. Undifferentiated follicular, papillary, and aggressive anaplastic carcinomas that showed poor prognosis and strong tumor invasive and metastatic potential and that can be compared in their morphological, genetic and growth features with undifferentiated thyroid carcinoma cell lines were not included this

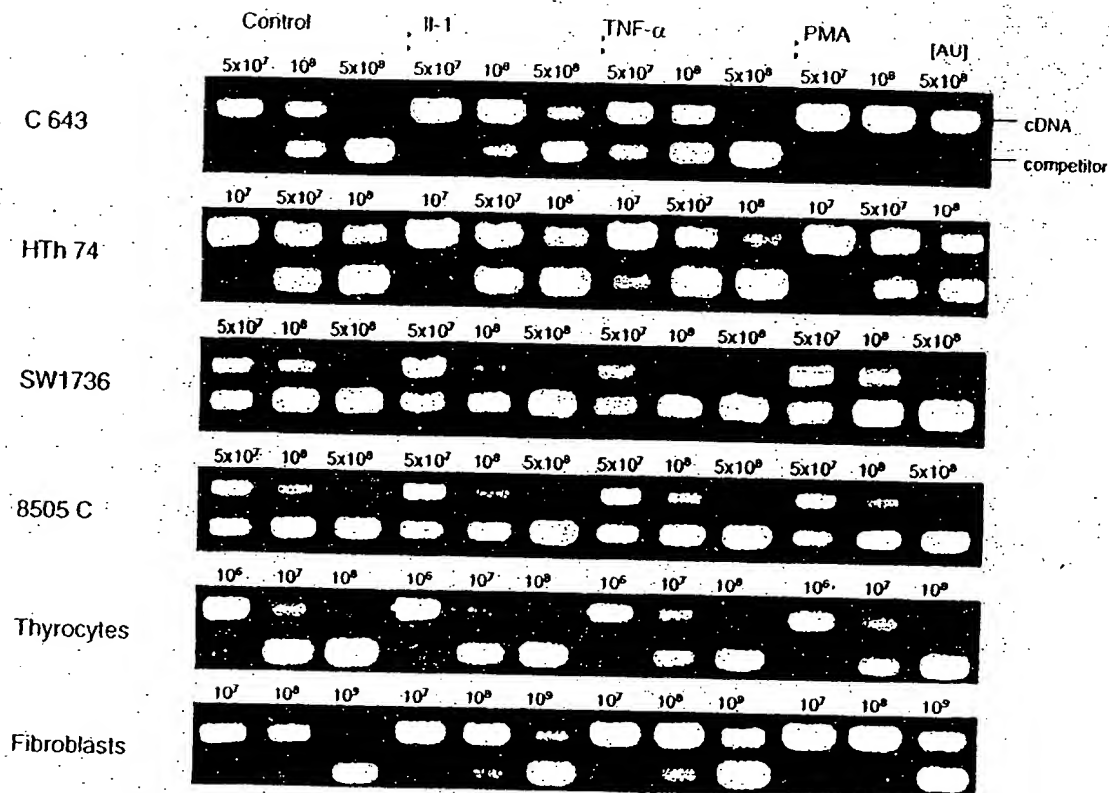


FIG. 4. Competitive TIMP-1 mRNA RT-PCR yielding a 351-bp (cDNA) and a 274-bp (competitor) PCR product. For further details see Figure 3.

study. However, the missing expression of MMP-1 by normal thyrocytes and the spontaneous secretion of this protein by highly malignant thyroid carcinoma cell lines, as demonstrated in our study, indicate the involvement of MMP-1 secretion of transformed thyrocytes in aggressive thyroid tumors.

Although all cell lines analyzed in our study spontaneously secreted MMP-1, we observed marked differences in the basal secretion capacity. The highest MMP-1 levels were determined in cultures of 8505 C cells. Only 8505 C cells expressed TPO and Tg mRNA that may be put down to residual differentiated components in the cell line (see *Materials*). However, none of the analyzed cell lines expressed TSH-R mRNA. The cell population doubling times were less than 40 hours. All cell lines had accumulations of multiple genetic events. These facts indicate the undifferentiated pathology of the studied lines. It is well known that anaplastic carcinoma cell lines well retain the malignant characteristics of their parental tumors (38–40).

Furthermore, we found a distorted proportion between MMP-1 and TIMP-1 mRNA/protein for carcinoma cell lines but not for normal thyroid-derived fibroblasts. The most disadvantageous constellation between MMP-1 and TIMP-1 was found in 8505 C cells. Similar to other studies (41), these results suggest the influence of an altered MMP/TIMP relation on tumor progression. However, it

should be mentioned that most studies, including the present one, do not take into consideration that a number of inhibitors distinct from TIMP-1 may regulate MMP-1 activity. Taking into account that the balance of active enzyme and TIMP-1 concentration strongly influence the extent of local matrix degradation, a number of studies showed unexpectedly high levels of TIMP-1 in malignant neoplasms (9,42,43). There is a great discussion as to whether the overall expression of MMP-1 and TIMP-1 or the amount of noncomplexed MMP-1 could be critical in aggressive tumor development. This fact underlines the nature of tissue breakdown, reflecting the complicated network of selective and coordinated production of individual proteinases and inhibitors under normal and pathophysiological conditions. Thus, the invasive and metastatic potential of thyroid tumors depends on the local net level of active MMPs.

The synthesis of MMP-1 and TIMP-1 is influenced by a variety of biochemical stimuli. The recent findings on MMP-1 and TIMP-1 gene promoters are useful in understanding the complex mechanisms implied in the regulation of MMP synthesis modulated by cytokines and tumor promoters (34,44,45). The promoter regions contain tumor promoter responsive elements (TRE) and binding motifs for the transcription factor PEA-3, which are recognized by proto-oncogenic transcription factors, such as the



FIG. 5. MMP-1 protein levels (mean  $\pm$  SEM) in supernatants of unstimulated and stimulated cultures of fibroblasts ( $n = 3$ ), thyroid-derived fibroblasts ( $n = 4$ ) and the thyroid carcinoma cell lines 8505 C, SW 1736, C 643, HTh 74 ( $n = 3$ ) detected by MMP-1 ELISA, which recognizes total MMP-1, i.e., free MMP-1 and that complexed with inhibitors such as TIMP-1, but not  $\alpha_2$ -macroglobulin. Cells were stimulated with 10 U/mL IL-1 $\alpha$ , 100 U/mL TNF- $\alpha$ , 500 U/mL IFN- $\gamma$ , or 10 ng/mL PMA for 24 and 48 hours. For fibroblasts, each point represents the mean  $\pm$  SEM of four different donors, each experiment performed in triplicate. For cell lines, each point represents the mean  $\pm$  SEM of three separate experiments each performed in triplicate. Significant differences between the basal and stimulated MMP-1 levels are indicated by asterisks (\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .005$ ). Please note the differences in scale between 8505 C, HTh 74, and the other cell types.

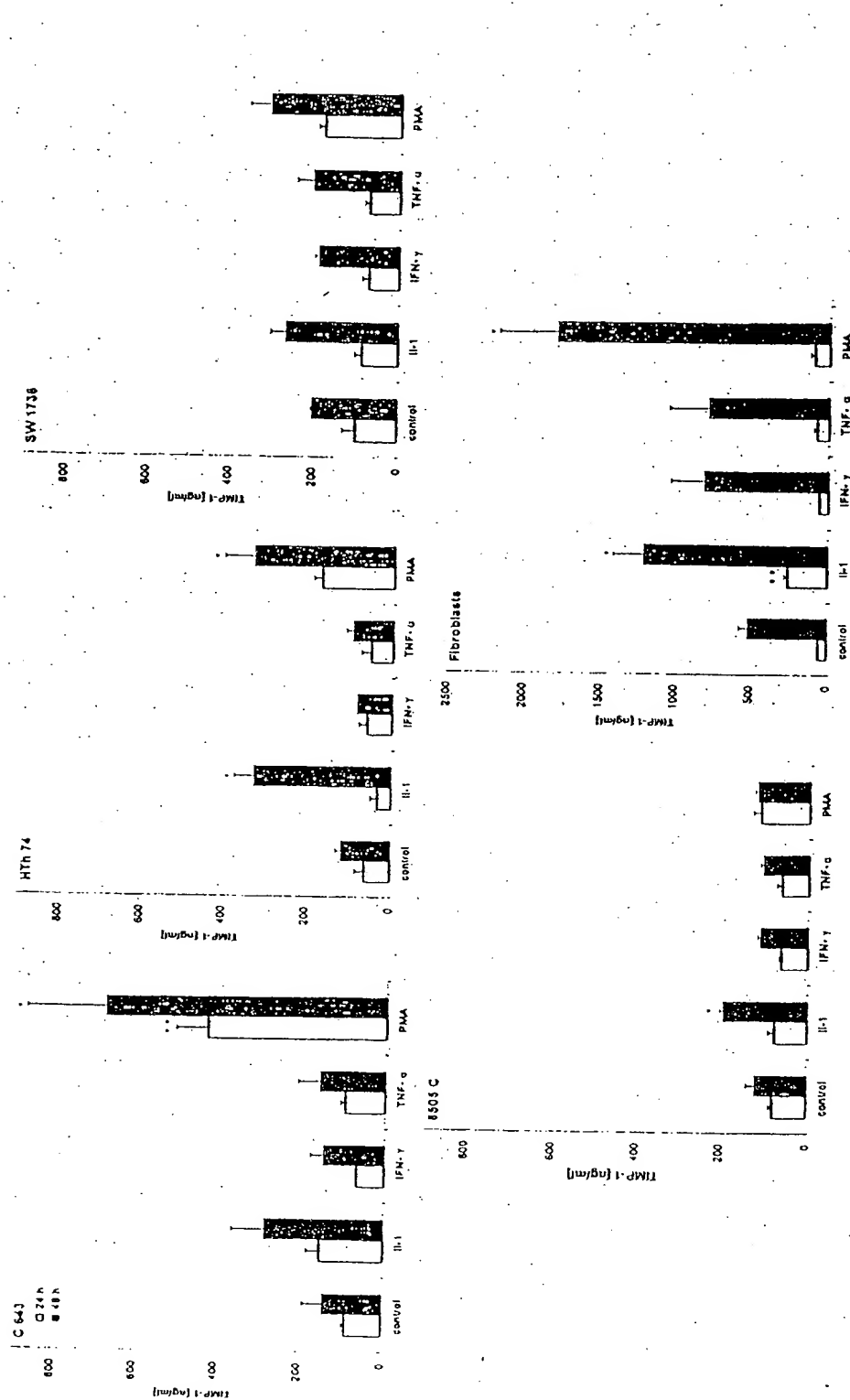


FIG. 6. ELISA detection of free and MMP-complexed TIMP-1 (mean  $\pm$  SEM). Asterisks indicate significant differences between the basal and stimulated TIMP-1 levels ( $*p < 0.05$ ;  $**p < 0.01$ ;  $***p < 0.005$ ). For further details see Figure 5. Please note the differences in scale between fibroblasts and the other cell types.

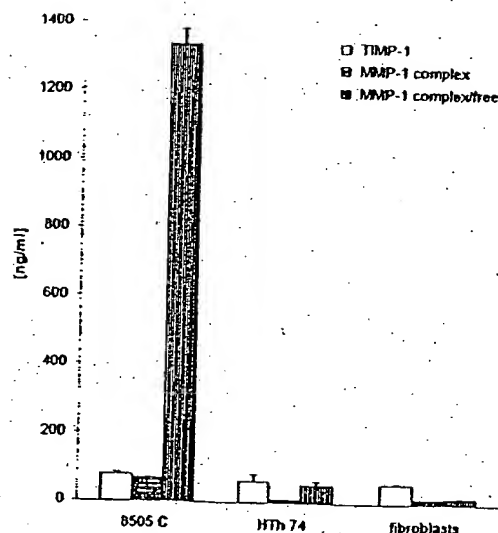


FIG. 7. Comparison between (i) free/complex and (ii) TIMP-1 complexed MMP-1 levels, and (iii) TIMP-1 levels in supernatants of unstimulated 8505 C and HTh 74 cells, and thyroid-derived fibroblasts after 24 hours using: a (i) MMP-1 ELISA that recognizes total MMP-1 (see Figure 5). The (ii) MMP-1/TIMP-1 assay recognizes MMP-1/TIMP-1 complexes, ie, activated MMP-1 that has subsequently been complexed with the specific MMP inhibitor TIMP-1. The (iii) TIMP-1 ELISA recognizes total TIMP-1, ie, free TIMP-1 and that complexed with MMPs.

*fos* and *jun* family (45-47). IL-1, TNF- $\alpha$ , and PMA up-regulate proto-oncogenes like *fos* and *jun*, resulting in the stimulation of MMP-1 and TIMP-1 (45,48). The action of the cytokines is mediated by their specific receptors. In our study, IL-1R (type I and type II), TNF- $\alpha$ R (p75 and p55) and IFN- $\gamma$ R mRNAs were demonstrated in all investigated cell types. PMA and IL-1 were shown to elevate MMP-1 and TIMP-1 in nearly all cell types investigated, thus confirming the results of several studies on other epithelial cells (reviewed in refs. 1,7,49). In the majority of experiments, we found a concordant expression of MMP-1 and TIMP-1 after stimulation, possibly achieved by the coordinated actions of the nuclear transcription factors, although MMP-1 and TIMP-1 expression can also be independently or even reciprocally regulated (1). The effect of TNF- $\alpha$  was not as distinct as in the case of PMA and IL-1, although several investigators found a pronounced effect of TNF- $\alpha$  particularly on TIMP-1 secretion (4,34). In contrast to studies performed with other cell types (863,864,819), IFN- $\gamma$  did not influence MMP-1 and TIMP-1 expression in thyroid carcinoma cell lines. In summary, the involvement of the intrathyroidal physiological and pathological cytokine microenvironment in the regulation of MMP-1 and TIMP-1 induction activation and inhibition is strongly suggested.

Furthermore, the data demonstrate that regular human thyrocytes did not produce MMP-1, even after powerful stimulation with PMA. Investigating other mammalian epithelial cells, only one study revealed the production of

MMP-1 by rabbit corneal cells (21). It is yet not clear whether the MMP-1 mRNA detected in thyrocytes is due to a low level of constitutive transcription of the MMP-1 gene (illegitimate transcription), an existing pool of stable MMP-1 mRNA, or *in vitro* induction of MMP-1 mRNA. But it seems more likely that residual fibroblasts contained in the purified thyrocyte preparation (<0.2%) are responsible for the slightly positive RT-PCR results. Another explanation could be that thyrocytes are indeed MMP-1 producers, but the ELISA detection system used was not sensitive enough to measure extremely low MMP-1 secretion levels. Furthermore, the discrepancy between elevated TIMP-1 mRNA levels of thyrocytes and the extremely low TIMP-1 protein secretion by these cells is difficult to explain. Post-transcriptional regulatory events may be responsible for this confounding result.

Taken together, the present study suggests that the intrathyroidal cytokine microenvironment is involved in the regulation of MMP-1 and its inhibitor TIMP-1 in the thyroid, and that both proteins may be secreted by dedifferentiated thyroid carcinoma cells and involved in aggressive thyroid tumors *in vivo*.

## REFERENCES

1. Ries C, Petrides PE 1995 Cytokine regulation of matrix metalloproteinase activity and its regulatory dysfunction in disease. *Biol Chem Hoppe Seyler* 376:345-355.
2. Pierce RA, Sandefur S, Doyle GA, Welgus HG 1996 Monocytic cell type-specific transcriptional induction of collagenase. *J Clin Invest* 97:1890-1899.
3. Pan L, Eckhoff C, Brinckerhoff CE 1995 Suppression of collagenase gene expression by all-trans and 9-cis retinoic acid is ligand dependent and requires both RARs and RXRs. *J Cell Biochem* 57:575-589.
4. Shingu M, Nagai Y, Isayama T, Naono T, Nobunaga M 1993 The effects of cytokines on metalloproteinase inhibitors (TIMP) and collagenase production by human chondrocytes and timp production by synovial cells and endothelial cells. *Clin Exp Immunol* 94:145-149.
5. Goldberg GI, Wilhelm SM, Kronberger A, Bauer EA, Grant GA, Eisen AZ 1986 Human fibroblast collagenase. Complete primary structure and homology to an oncogene transformation-induced rat protein. *J Biol Chem* 261:6600-6605.
6. Docherty AJ, Lyons A, Smith BJ, Wright EM, Stephens PE, Harris TJ, Murphy G, Reynolds JJ 1985 Sequence of human tissue inhibitor of metalloproteinases and its identity to erythroid-potentiating activity. *Nature* 318:66-69.
7. Mauviel A 1993 Cytokine regulation of metalloproteinase gene expression. *J Cell Biochem* 53:288-295.
8. Murphy G 1995 Matrix metalloproteinases and their inhibitors. *Acta Orthop Scand Suppl* 266:55-60.
9. Nomura H, Fujimoto N, Seiki M, Mai M, Okada Y 1996 Enhanced production of matrix metalloproteinases and activation of matrix metalloproteinase 2 (gelatinase A) in human gastric carcinomas. *Int J Cancer* 69:9-16.
10. Murray GI, Duncan ME, O'Neil P, Melvin WT, Fothergill JE 1996 Matrix metalloproteinase-1 is associated with poor prognosis in colorectal cancer. *Nat Med* 2:461-462.
11. Onisto M, Garbisa S, Caenazzo C, Freda MP, Di Francesco C, Nitti D, Liotta LA, Stetler-Stevenson WG 1993 Reverse transcription-polymerase chain reaction phenotyping of met-

- allopeptinases and inhibitors involved in tumor matrix invasion. *Diagn Mol Pathol* 2:74-80.
12. Kameyama K 1996 Expression of MMP-1 in the capsule of thyroid cancer—Relationship with invasiveness. *Pathol Res Pract* 192:20-26.
13. Campo E, Merino MJ, Liotta L, Neumann R, Stetler-Stevenson W 1992 Distribution of the 72-kd type IV collagenase in nonneoplastic and neoplastic thyroid tissue. *Hum Pathol* 23:1395-1401.
14. Demeure MJ, Damsky CH, Elfman F, Coretzki PE, Wong MG, Clark O 1992 Invasion by cultured human follicular thyroid cancer correlates with increased beta 1 integrins and production of proteases. *World J Surg* 16:770-776.
15. Nakano T, Kusunoki T, Funasaka K, Murata K, Nishida S, Tomura T 1995 Study of type I and IV collagenase activity in human thyroid diseases. *Nippon Jibiinkoka Gakkai Kaiho* 98:937-941.
16. Zedenius J, Stahle-Backdahl M, Enberg U, Grmelius L, Larsson C, Wallin G, Backdahl M 1996 Stromal fibroblasts adjacent to invasive thyroid tumors: expression of gelatinase A but not stromelysin 3 mRNA. *World J Surg* 20:101-106.
17. Arai M, Niioka M, Maruyama K, Wada N, Fujimoto N, Nomiyama T, Tanaka S, Okazaki I 1996 Changes in serum levels of metalloproteinases and their inhibitors by treatment of chronic hepatitis C with interferon. *Dig Dis Sci* 41:995-1000.
18. Nikkari ST, O'Brien KD, Ferguson M, Hatsukami T, Welgus HG, Alpers CE, Clowes AW 1995 Interstitial collagenase (MMP-1) expression in human carotid atherosclerosis. *Circulation* 92:1393-1398.
19. Knowlden J, Martin J, Davies M, Williams JD 1995 Metalloproteinase generation by human glomerular epithelial cells. *Kidney Int* 47:1682-1689.
20. Buisson AC, Zahm JM, Polette M, Pierrot D, Bellon G, Puchelle E, Birembaut P, Tournier JM 1996 Gelatinase B is involved in the in vitro wound repair of human respiratory epithelium. *J Cell Physiol* 166:413-426.
21. Tao Y, Bazan HE, Bazan NG 1995 Platelet-activating factor induces the expression of metalloproteinases-1 and -9, but not -2 or -3, in the corneal epithelium. *Invest Ophthalmol Vis Sci* 36:345-354.
22. Aust G, Heuer M, Laue S, Lehmann I, Hofmann A, Heldin N-E, Scherbaum WA 1996 Expression of TNA-alpha mRNA and protein in pathological thyroid tissue and carcinoma cell lines. *Clin Exp Immunol* 105:148-154.
23. Mark J, Ekedahl C, Dahlenfors R, Westermark B 1987 Cytogenetical observations in five human anaplastic thyroid carcinomas. *Hereditas* 107:163-174.
24. Heldin NE, Cvejic D, Smeds S, Westermark B 1991 Coexpression of functionally active receptors for thyrotropin and platelet-derived growth factor in human thyroid carcinoma cells. *Endocrinology* 129:2187-2193.
25. Ito T, Seyama T, Hayashi Y, Hayashi T, Dohi K, Mizuno T, Iwamoto KS, Tsuyama N, Nakamura N, Akiyama M 1994 Establishment of two human thyroid carcinoma cell lines (8305C, 8505C) bearing p53 gene mutations. *International Journal of Oncology* 4:583-586.
26. Wallin G, Backdahl M, Tallroth-Ekman E, Lundell G, Auer G, Lowhagen T 1989 Co-existent anaplastic and well differentiated thyroid carcinomas: a nuclear DNA study. *Eur J Surg Oncol* 15:43-48.
27. Heuer M, Aust G, Ode-Hakim S, Scherbaum WA 1996 Different cytokine mRNA profiles in Graves' disease, Hashimoto's thyroiditis and non-autoimmune thyroid disorders determined by quantitative reverse transcriptase chain reaction (RT-PCR). *Thyroid* 6:97-106.
28. Celi FS, Zenilman ME, Shuldiner AR 1993 A rapid and versatile method to synthesize internal standards for competitive PCR. *Nucleic Acids Res* 21:1047.
29. Platzer C, Ode-Hakim S, Reinke P, Docke WD, Ewert R, Volk HD 1994 Quantitative PCR analysis of cytokine transcription patterns in peripheral mononuclear cells after anti-CD3 rejection therapy using two novel multispecific competitor fragments. *Transplantation* 58:264-268.
30. Watson PF, Pickerill AP, Davies R, Weetman AP 1994 Analysis of cytokine gene expression in Graves' disease and multinodular goiter. *J Clin Endocrinol Metab* 79:355-360.
31. Tada M, Diserens AC, Desbaillets I, de Tribolet N 1994 Analysis of cytokine receptor messenger RNA expression in human glioblastoma cells and normal astrocytes by reverse-transcription polymerase chain reaction. *J Neurosurg* 80:1063-1073.
32. Nakano A, Tani E, Miyazaki K, Yamamoto Y, Furuyama J 1995 Matrix metalloproteinases and tissue inhibitors of metalloproteinases in human gliomas. *J Neurosurg* 83:298-307.
33. Whitelock JM, O'Grady RL, Gibbins JR 1991 Interstitial collagenase (matrix metalloproteinase 1) associated with the plasma membrane of both neoplastic and nonneoplastic cells. *Invasion Metastasis* 11:139-148.
34. Mackay AR, Ballin M, Pelina MD, Farina AR, Nason AM, Hartzler JL, Thorgeirsson UP 1992 Effect of phorbol ester and cytokines on matrix metalloproteinase and tissue inhibitor of metalloproteinase expression in tumor and normal cell lines. *Invasion Metastasis* 12:168-184.
35. Urbanski SJ, Edwards DR, Maitland A, Leco KJ, Watson A, Kossakowska AE 1992 Expression of metalloproteinases and their inhibitors in primary pulmonary carcinomas. *Br J Cancer* 66:1188-1194.
36. Polette M, Clavel C, Muller D, Abecassis J, Binniger I, Birembaut P 1991 Detection of mRNAs encoding collagenase 1 and stromelysin 2 in carcinomas of the head and neck by in situ hybridization. *Invasion Metastasis* 11:76-83.
37. Clavel C, Polette M, Doco M, Binniger I, Birembaut P 1992 Immunolocalization of matrix metalloproteinases and their tissue inhibitor in human mammary pathology. *Bull Cancer (Paris)* 79:261-270.
38. Boghaert ER, Ain K, Taylor K, Greenberg VL, Fowler C, Zimmer SG 1996 Quantitative and qualitative differences in growth, invasion and lung colonization of an anaplastic and a papillary human thyroid cancer cell line in vitro and in vivo. *Clin Exp Metastasis* 14:440-450.
39. Asakawa H, Kobayashi T, Komoike Y, Yamagawa T, Takahashi M, Wakasugi E, Maruyama H, Tamaki Y, Matsuzawa Y, Monden M 1996 Establishment of anaplastic thyroid carcinoma cell lines useful for analysis of chemosensitivity and carcinogenesis. *J Clin Endocrinol Metab* 81:3547-3552.
40. Viglietto G, Maglione D, Rambaldi M, Cerutti J, Romano A, Trapasso F, Fedele M, Ippolito P, Chiappetta G, Boti G, et al 1995 Upregulation of vascular endothelial growth factor (VEGF) and downregulation of placenta growth factor (PIGF) associated with malignancy in human thyroid tumors and cell lines. *Oncogene* 11:1569-1579.
41. Nuovo GJ, MacConnell PB, Simsir A, Valea F, French DL 1995 Correlation of the in situ detection of polymerase chain reaction-amplified metalloproteinase complementary dnas and their inhibitors with prognosis in cervical carcinoma. *Cancer Res* 55:267-275.
42. Naruo S, Kanayama H, Takigawa H, Kagawa S, Yamashita K, Hayakawa T 1994 Serum levels of a tissue inhibitor of

- metalloproteinases-1 (TIMP-1) in bladder cancer patients. *Int J Urol* 1:228-231.
43. Kossakowska AE, Urbanski SJ, Watson A, Hayden LJ, Edwards DR 1993 Patterns of expression of metalloproteinases and their inhibitors in human malignant lymphomas. *Oncol Res* 5:19-28.
44. Edwards DR, Rocheleau H, Sharma RR, Wills AJ, Cowie A, Hassell JA, Heath JK 1992 Involvement of AP1 and PEA3 binding sites in the regulation of murine tissue inhibitor of metalloproteinases-1 (TIMP-1) transcription. *Biochim Biophys Acta* 1171:41-55.
45. Schonthal A, Herrlich P, Rahmsdorf HJ, Ponta H 1988 Requirement for fos gene expression in the transcriptional activation of collagenase by other oncogenes and phorbol esters. *Cell* 54:325-334.
46. Gutman A, Wasylyk B 1990 The collagenase gene promoter contains a TPA and oncogene responsive unit encompassing the PEA3 and AP-1 binding sites. *EMBO J* 9:2241-2246.
47. Wasylyk C, Gutman A, Nicholson R, Wasylyk B 1991 The c-Ets oncoprotein activates the stromelysin promoter through the same elements as several non-nuclear oncoproteins. *EMBO J* 10:1127-1134.
48. Brenner DA, O'Hara M, Angel P, Chojkier M, Karin M 1989 Prolonged activation of jun and collagenase genes by tumour necrosis factor-alpha. *Nature* 337:661-663.
49. Opdenakker G, Van Damme J 1992 Cytokines and proteases in invasive processes: molecular similarities between inflammation and cancer. *Cytokine* 4:251-258.

Address reprint requests to:  
Dr. Gabriela Aust  
Institute of Anatomy  
University of Leipzig  
Liebigstr. 13  
Leipzig, D-04103, Germany



**Expression of embryonic fibronectin isoform EIIIA parallels alpha-smooth muscle actin in maturing and diseased kidney.**

**Barnes VL, Musa J, Mitchell RJ, Barnes JL.**

Department of Medicine, Division of Nephrology, University of Texas Health Science Center, San Antonio, Texas, USA.

In this study we examined if an association exists between expression of an alternatively spliced "embryonic" fibronectin isoform EIIIA (Fn-EIIIA) and alpha-smooth muscle actin (alpha-SMA) in the maturing and adult rat kidney and in two unrelated models of glomerular disease, passive accelerated anti-glomerular basement membrane (GBM) nephritis and Habu venom (HV)-induced proliferative glomerulonephritis, using immunohistochemistry and in situ hybridization. Fn-EIIIA and alpha-SMA proteins were abundantly expressed in mesangium and in periglomerular and peritubular interstitium of 20-day embryonic and 7-day (D-7) postnatal kidneys in regions of tubule and glomerular development. Staining was markedly reduced in these structures in maturing juvenile (D-14) kidney and was largely lost in adult kidney. Expression of Fn-EIIIA and alpha-SMA was reinitiated in the mesangium and the periglomerular and peritubular interstitium in both models and was also observed in glomerular crescents in anti-GBM nephritis. Increased expression of Fn-EIIIA mRNA by in situ hybridization corresponded to the localization of protein staining. Dual labeling experiments verified co-localization of Fn-EIIIA and alpha-SMA, showing a strong correlation of staining between location and staining intensity during kidney development, maturation, and disease. Expression of EIIIA mRNA corresponded to protein expression in developing and diseased kidneys and was lost in adult kidney. These studies show a recapitulation of the co-expression of Fn-EIIIA and alpha-SMA in anti-GBM disease and suggest a functional link for these two proteins.

PMID: 10330455 [PubMed - indexed for MEDLINE]



Comment in:

- [Cancer Res. 2002 Jan 15;62\(2\):618-9.](#)

[Free full text article at  
cancerres.aacrjournals.org](#)

**BMI-1 gene amplification and overexpression in hematological malignancies occur mainly in mantle cell lymphomas.**

Bea S, Tort F, Pinyol M, Puig X, Hernandez L, Hernandez S, Fernandez PL, van Lohuizen M, Colomer D, Campo E.

The Hematopathology Section, Laboratory of Anatomic Pathology, Hospital Clinic, Institut d'Investigacions Biomediques August Pi i Sunyer (IDIBAPS), University of Barcelona, Spain.

The BMI-1 gene is a putative oncogene belonging to the Polycomb group family that cooperates with c-myc in the generation of mouse lymphomas and seems to participate in cell cycle regulation and senescence by acting as a transcriptional repressor of the INK4a/ARF locus. The BMI-1 gene has been located on chromosome 10p13, a region involved in chromosomal translocations in infant leukemias, and amplified in occasional non-Hodgkin's lymphomas (NHLs) and solid tumors. To determine the possible alterations of this gene in human malignancies, we have examined 160 lymphoproliferative disorders, 13 myeloid leukemias, and 89 carcinomas by Southern blot analysis and detected BMI-1 gene amplification (3- to 7-fold) in 4 of 36 (11%) mantle cell lymphomas (MCLs) with no alterations in the INK4a/ARF locus. BMI-1 and p16INK4a mRNA and protein expression were also studied by real-time quantitative reverse transcription-PCR and Western blot, respectively, in a subset of NHLs. BMI-1 expression was significantly higher in chronic lymphocytic leukemia and MCL than in follicular lymphoma and large B cell lymphoma. The four tumors with gene amplification showed significantly higher mRNA levels than other MCLs and NHLs with the BMI-1 gene in germline configuration. Five additional MCLs also showed very high mRNA levels without gene amplification. A good correlation between BMI-1 mRNA levels and protein expression was observed in all types of lymphomas. No relationship was detected between BMI-1 and p16INK4a mRNA levels. These findings suggest that BMI-1 gene alterations in human neoplasms are uncommon, but they may contribute to the pathogenesis in a subset of malignant lymphomas, particularly of mantle cell type.

PMID: 11289106 [PubMed - indexed for MEDLINE]

## BMI-1 Gene Amplification and Overexpression in Hematological Malignancies Occur Mainly in Mantle Cell Lymphomas<sup>1</sup>

Silvia Bea, Frederic Tort, Magda Pinyol, Xavier Puig, Luis Hernández, Silvia Hernández, Pedro L. Fernández, Maarten van Lohuizen, Dolors Colomer, and Elias Campo<sup>2</sup>

The Hematopathology Section, Laboratory of Anatomic Pathology, Hospital Clinic, Institut d'Investigacions Biomèdiques "August Pi i Suñer" (IDIBAPS), University of Barcelona, 08036 Barcelona, Spain [S. B., F. T., M. P., X. P., L. H., S. H., P. L. F., D. C., E. C.], and Division of Molecular Carcinogenesis, The Netherlands Cancer Institute, 1066 CX Amsterdam, Netherlands [M. v. L.]

### Abstract

The *BMI-1* gene is a putative oncogene belonging to the Polycomb group family that cooperates with *c-myc* in the generation of mouse lymphomas and seems to participate in cell cycle regulation and senescence by acting as a transcriptional repressor of the *INK4a/ARF* locus. The *BMI-1* gene has been located on chromosome 10p13, a region involved in chromosomal translocations in infant leukemias, and amplified in occasional non-Hodgkin's lymphomas (NHLs) and solid tumors. To determine the possible alterations of this gene in human malignancies, we have examined 160 lymphoproliferative disorders, 13 myeloid leukemias, and 89 carcinomas by Southern blot analysis and detected *BMI-1* gene amplification (3- to 7-fold) in 4 of 36 (11%) mantle cell lymphomas (MCLs) with no alterations in the *INK4a/ARF* locus. *BMI-1* and *p16<sup>INK4a</sup>* mRNA and protein expression were also studied by real-time quantitative reverse transcription-PCR and Western blot, respectively, in a subset of NHLs. *BMI-1* expression was significantly higher in chronic lymphocytic leukemia and MCL than in follicular lymphoma and large B cell lymphoma. The four tumors with gene amplification showed significantly higher mRNA levels than other MCLs and NHLs with the *BMI-1* gene in germline configuration. Five additional MCLs also showed very high mRNA levels without gene amplification. A good correlation between *BMI-1* mRNA levels and protein expression was observed in all types of lymphomas. No relationship was detected between *BMI-1* and *p16<sup>INK4a</sup>* mRNA levels. These findings suggest that *BMI-1* gene alterations in human neoplasms are uncommon, but they may contribute to the pathogenesis in a subset of malignant lymphomas, particularly of mantle cell type.

### Introduction

The *BMI-1*<sup>3</sup> gene is a putative oncogene of the Polycomb group originally identified by retroviral insertional mutagenesis in Eμ-*c-myc* transgenic mice infected with the Moloney murine leukemia virus (1, 2). These animals had a rapid development of pre-B cell lymphomas showing frequent proviral insertions near the *BMI-1* gene. This integration resulted in *BMI-1* overexpression suggesting a cooperative effect between *C-MYC* and *BMI-1* genes in the development of these tumors (3, 4). Recent studies have indicated that the *BMI-1* gene may also participate in cell cycle control and senescence through the

*INK4a/ARF* locus by acting as an upstream negative regulator of *p16<sup>INK4a</sup>* and *p14/p19<sup>ARF</sup>* gene expression (5). The human *BMI-1* gene has been mapped to chromosome 10p13 (6), a region involved in chromosomal translocations in infant leukemias (7) and rearrangements in malignant T cell lymphomas (8, 9). More recently, high-level DNA amplifications of this region have been found by comparative genomic hybridization in NHLs and solid tumors (10, 11). However, the possible implication of the *BMI-1* gene in these alterations and its role in the pathogenesis of human tumors is not known. The aim of this study was to analyze the possible *BMI-1* gene alterations and expression in a large series of human neoplasms and to determine the relationship with *INK4a/ARF* locus aberrations.

### Materials and Methods

**Case Selection.** A series of 262 human tumors, including 173 hematological malignancies and 89 carcinomas (Table 1), matched normal tissues from all carcinomas, 11 samples of normal peripheral mononuclear cells, and 5 reactive lymph nodes and tonsils, were selected based on the availability of frozen samples for molecular analysis.

**DNA Extraction and Southern Blot Analysis.** Genomic DNA was obtained using Proteinase K/RNase treatment. 15 μg were digested with *EcoRI* and *HindIII* restriction enzymes (Life Technologies, Inc., Gaithersburg, MD), for Southern blot analysis and hybridized with a 1.5-kb *PstI* fragment of the partial *BMI-1* cDNA (6).

**RNA Extraction and Real-time Quantitative RT-PCR.** Total RNA was obtained from 67 lymphoid neoplasms (10 CLLs, 27 MCLs, 8 FLs, and 22 LCLs) using guanidine/isothiocyanate extraction and cesium/chloride gradient centrifugation. One μg of total RNA was transcribed into cDNA using MMLV-reverse transcriptase (Life Technologies, Inc.) and random hexamers, following manufacturer's directions. Sequences of the *BMI-1* and the *p16* detection probes and primers were designed using the Primer Express program (Applied Biosystems, Foster City) as follows: *BMI-1* sense, 5'-CTGGTTGCCATTGACAGC-3'; *BMI-1* antisense, 5'-CAGAAAATGAATGCGAGCCA-3'; *p16* sense, 5'-CAACGCACCGAATAGTTACGG-3'; *p16* antisense, 5'-AACTTCGCTCCAGAGTCGC-3'. The probes *BMI-1*, 5'-CAGCTCGCTCAAGATGGCCGC-3', and *p16*, 5'-CGGAGGCCGATCCAGGTGGTA-3', were labeled with 6-carboxy-fluorescein as the reporter dye. The TaqMan-GAPDH Control Reagents (Applied Biosystems) were used to amplify and detect the *GAPDH* gene, as recommended by the manufacturer. The quantitative assay amplified 1 μl of cDNA in two to four replicates using the primers and probes described above and the standard master mix (Applied Biosystems). All reactions were performed in an ABI PRISM 7700 Sequence Detector System (Applied Biosystems). *GAPDH*, *BMI-1*, and *p16<sup>INK4a</sup>* expression was related to a standard curve derived from serial dilutions of Raji cDNA. The RUs of *BMI-1* and *p16<sup>INK4a</sup>* expression were defined as the mRNA levels of these genes normalized to the *GAPDH* expression level in each case.

**Protein Analysis.** Whole-cell protein extracts were obtained from additional frozen tissue available in 31 cases (7 CLLs, 12 MCLs, 8 FLs, and 4 LCLs), loaded onto a 10% SDS-polyacrylamide gel, and electroblotted to a nitrocellulose membrane (Amersham). Blocked membranes were incubated sequentially with the monoclonal antibody BMI-F6 (12), antimuscle conju-

Received 10/16/00; accepted 1/29/01.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

<sup>1</sup> Supported by Grant SAF 99/20 from Comisión Interministerial de Ciencia y Tecnología, European Union Contract QLGI-CT-2000-689, the Asociación Española contra el Cáncer, and Generalitat de Catalunya 98SGR21. S. B. and F. T. were fellows supported by Spanish Ministerio de Educación y Cultura, and S. H. was supported by the Asociación Española contra el Cáncer and the Fundació Rius i Virgili.

<sup>2</sup> To whom requests for reprints should be addressed, at the Department of Pathology, Hospital Clinic, University of Barcelona, Villarroel 170, 08036 Barcelona, Spain. Phone: 34 93 227 5450; Fax: 34 93 227 5572; E-mail: campo@medicina.ub.es.

<sup>3</sup> The abbreviations used are: *BMI-1*, B cell-specific Moloney murine leukemia virus integration site 1; NHL, non-Hodgkin's lymphoma; CLL, chronic lymphocytic leukemia; FL, follicular lymphoma; LCL, large B cell lymphoma; MCL, mantle cell lymphoma; RT-PCR, reverse transcription-PCR; RU, relative units.

Table 1 Hematological malignancies and solid tumor samples analyzed for BMI-1 gene alterations

Tissue samples	No. of cases
Hematological malignancies	
Hodgkin's disease	2
B cell lymphoproliferative disorders	
B-Acute lymphoblastic leukemia	14
CLL	29
Hairy cell leukemia	4
FL	15
MCL	36
LCL	40
T cell lymphoproliferative disorders	
T-Acute lymphoblastic leukemia	8
Large granular cell leukemia	4
Peripheral T-cell lymphoma	8
Myeloproliferative disorders	
Acute myeloid leukemia	7
Chronic myeloid leukemia	6
Solid tumors	
Colon carcinoma	26
Breast carcinoma	29
Laryngeal squamous cell carcinoma	34
Total	262

gated to horseradish peroxidase (Amersham), and detected by enhanced chemiluminescence (Amersham) according to the manufacturer's recommendations.

**Statistical Analysis.** Because of the non-normal distribution of the samples and the small size of some subsets of tumors, the statistical evaluation was performed using nonparametric tests (SPSS, version 9.0). Comparison between mRNA expression levels in the different groups of NHLs was performed using the Kruskal-Wallis Test, with a *P* for significance set at 0.05. For differences between particular groups, the conservative Bonferroni procedure was performed, and the *P* was set at 0.005. The remaining statistical analyses were carried out using the Mann-Whitney nonparametric *U* test (significance, *P* < 0.05). The comparison between BMI-1 and p16<sup>INK4a</sup> quantitative mRNA levels was also performed using the Pearson's correlation coefficient.

## Results

**BMI-1 Gene Amplification.** The BMI-1 gene was examined by Southern blot in a large series of human tumors and normal samples (Table 1). The cDNA probe used in the study detected three *Eco*RI fragments of 7.3, 3.8, and 2.6 kb and three *Hind*III fragments of 6.2, 4, and 3.5 kb. BMI-1 gene amplification (3- to 7-fold) was detected in 4 of 36 (11%) MCLs (Fig. 1). The amplifications were confirmed with both restriction enzymes. The amplified MCLs were two blastoid and two typical variants. No amplifications were observed in any of the solid tumors when compared with their respective matched non-neoplastic mucosa. No BMI-1 gene rearrangements were observed in any of the samples examined.

**BMI-1 mRNA Expression.** To determine the BMI-1 expression pattern in NHL we analyzed BMI-1 mRNA levels by real-time quantitative RT-PCR in 67 lymphomas (10 CLLs, 27 MCLs, 8 FLs, and 22 LCLs), including the four tumors with gene amplification. A distinct BMI-1 mRNA expression pattern was observed in the different types of lymphomas (Fig. 2; Kruskal-Wallis Test; *P* < 0.001). The BMI mRNA levels in CLLs (mean, 2.2 RU; SD, 1.3) and MCLs with no BMI-1 gene amplification (mean, 2.5 RU; SD, 2.3) were significantly higher than in FLs (mean, 0.9 RU; SD, 0.8) and LCLs (mean, 0.6 RU; SD, 0.4; Mann-Whitney nonparametric *U* test; *P* < 0.01). The 4 MCLs with BMI-1 gene amplification showed significantly higher levels of expression than all other groups of tumors (mean, 5.1 RU; SD, 1.6; *P* < 0.005). In addition, five typical MCLs with no structural alterations of the gene also showed very high levels of BMI-1 mRNA expression ranging from 4 to 9.8 RU, similar to cases with gene amplification (Fig. 2A).

**BMI-1 Protein Expression.** BMI-1 protein expression was examined by Western blot in 31 tumors (7 CLLs; 12 MCLs, including two

cases with BMI-1 gene amplification and 4 cases with mRNA overexpression and no structural alteration of the gene; 8 FLs, and 4 LCLs) in which additional frozen tissue was available. The monoclonal antibody against BMI-1 detected three closely migrating proteins of *M<sub>r</sub>* 45,000–48,000 (2). The two more slowly migrating bands probably represent phosphorylated isoforms of the protein (12). The two MCLs with gene amplification and three of four cases with mRNA overexpression without amplification of the gene showed very high levels of protein expression. The remaining MCLs and CLLs showed intermediate levels of protein expression, whereas low- or no-expression signals were detected in the LCLs and FLs included in the study (Fig. 3). These results indicate that BMI-1 protein expression in NHL is concordant with the mRNA levels observed by real-time quantitative RT-PCR.

**Relationship between BMI-1 and p16<sup>INK4a</sup> Gene Alterations.** The *INK4a/ARF* locus has been recently identified as a downstream target of the transcriptional repressing activity of the BMI-1 gene, suggesting that this gene may contribute to human neoplasias with wild type *INK4a/ARF* (5). Most of the lymphoproliferative disorders analyzed in the present study, including the four cases with BMI-1 gene amplification, had been previously examined for p53 gene mutations and *INK4a/ARF* locus alterations, including gene deletions, mutations, hypermethylation, and expression (13, 14). The four MCLs with BMI-1 gene amplification and mRNA overexpression and the five tumors with BMI-1 mRNA overexpression with no structural alterations of the gene showed a wild-type configuration of the *INK4a/ARF* locus (13). However, one case with BMI-1 gene amplification and one case with mRNA overexpression with no alteration of the gene showed p53 gene mutations associated with allelic deletions.

To determine the possible relationship between BMI-1 and p16<sup>INK4a</sup> mRNA expression, p16<sup>INK4a</sup> mRNA levels were evaluated by real-time quantitative RT-PCR in 50 tumors (10 CLLs, 27 MCLs, and 13 LCLs), including 6 cases with alterations in the *INK4a/ARF* locus (2 MCLs and 1 LCL with p16<sup>INK4a</sup> gene deletion, 2 LCLs with p16 promoter hypermethylation, and 1 CLL with p16<sup>INK4a</sup> gene mutation), and the 4 lymphomas with BMI-1 amplification. Negative or negligible levels of p16<sup>INK4a</sup> were observed in the 6 tumors with *INK4a/ARF* locus alterations. These cases were not included in the comparisons between BMI-1 and p16<sup>INK4a</sup> mRNA expression. The p16<sup>INK4a</sup> expression levels were relatively similar in the different types of tumors. Only LCLs tended to have lower levels of expression, but the differences did not reach statistical significance (Fig. 2B). No differences were observed in the p16<sup>INK4a</sup> mRNA levels between tumors with BMI-1 gene amplification and overexpression and lymphomas with germline configuration of the gene.

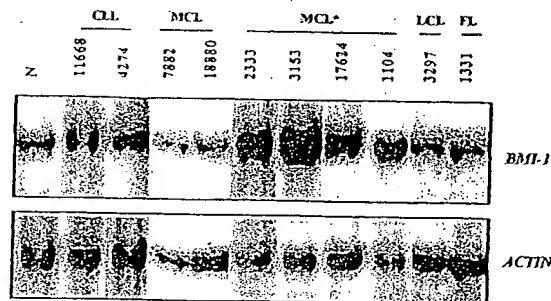


Fig. 1. Southern blot analysis of BMI-1 gene. Four MCLs (MCL\*) showed BMI-1 gene amplification (3- to 7-fold) compared with non-neoplastic tissues (N) and other NHLs. No amplifications or gene rearrangements were detected in the remaining NHLs and carcinomas included in the study.

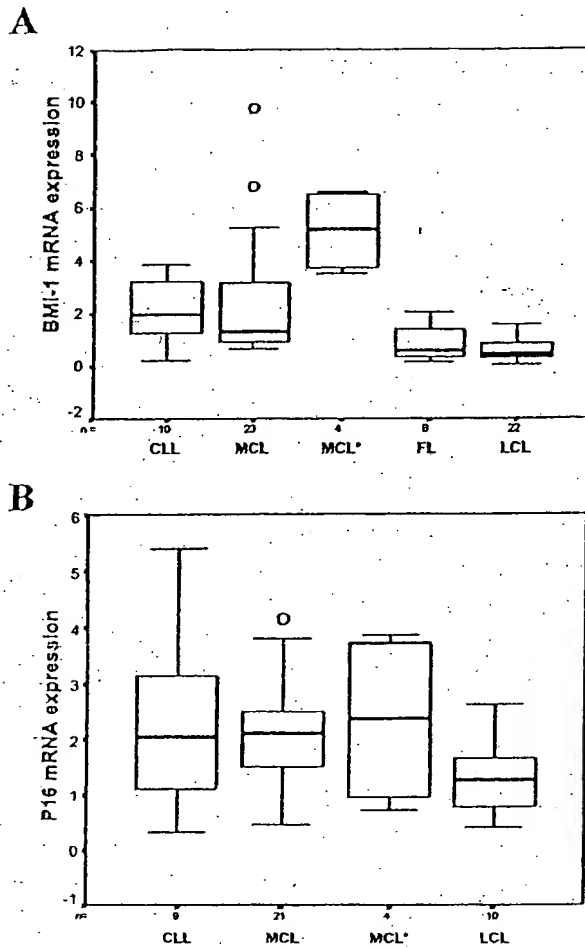


Fig. 2. A, quantitative BMI-1 mRNA transcript analysis (median and range) using real-time RT-PCR in a series of NHLs. MCLs with BMI-1 gene amplification (MCL\*) revealed significantly higher overall BMI-1 mRNA levels than all other types of NHLs, including MCLs with no structural alterations of the gene ( $P < 0.005$ ). MCLs and CLLs expressed significantly higher levels than FLs and LCLs ( $P < 0.001$ ). Results are depicted as the ratio of absolute BMI-1:GADPH mRNA transcript numbers (RU). Bars, SD. B, quantitative p16<sup>INK4a</sup> mRNA transcript analysis (median and range) using real-time RT-PCR in a series of NHLs. Expression levels were relatively similar in the different types of tumors. Results are depicted as the ratio of absolute p16<sup>INK4a</sup>:GADPH mRNA transcript numbers (RU). Bars, SD.

## Discussion

In the present study, we have examined a large series of human tumors for the presence of gene alterations and mRNA expression of the BMI-1 gene. Gene amplification was identified in four MCLs. These tumors showed significantly higher levels of mRNA and protein expression compared with other lymphomas with BMI-1 in germ-line configuration. BMI-1 expression levels were also highly up-regulated in a subset of MCLs with no apparent structural alterations of the gene. No alterations were detected in any of the different types of carcinomas included in the study. BMI-1 is considered an oncogene belonging to the Polycomb group family of genes. These proteins mainly act as transcriptional regulators, controlling specific target genes involved in development, cell differentiation, proliferation, and senescence. Different studies have shown the implication of BMI-1 overexpression in the development of lymphomas in murine and feline animal models (3, 4). The findings of the present study indicate

for the first time that BMI-1 gene alterations in human neoplasms are an uncommon phenomenon, but they seem to occur mainly in a subset of NHLs, particularly of mantle cell type.

The human BMI-1 gene has been mapped to chromosome 10p13. High-level DNA amplifications and gains in this region have been identified by comparative genomic hybridization in occasional solid tumors and NHLs (10, 11). Different chromosomal translocations involving the 10p13 region have also been identified in infant leukemias and T cell lymphoproliferative disorders (7, 8, 15). Most acute leukemias with this chromosomal alteration occur in children <12 months of age, whereas it seems to be extremely rare in adults. 10p translocations in T-cell lymphoproliferative disorders have been observed mainly in adult T cell leukemia/lymphomas and occasional cutaneous T cell lymphomas. In our study, we did not observe BMI-1 rearrangements or amplifications in any of the acute leukemias or T cell lymphomas. However, all of the acute leukemias in this study were diagnosed in patients over 16 years, and no adult T cell leukemia/lymphomas or cutaneous lymphomas could be included in the series. Similarly, high-level DNA amplifications at the 10p13 region have been detected in head and neck carcinomas and other solid tumors. Although we found no evidence for BMI-1 gene rearrangements or amplifications in a substantial set of carcinomas, this does not exclude the possibility of increased gene expression or protein levels in these tumors. Additional studies are required to elucidate the possible involvement of BMI-1 in these particular groups of human neoplasms.

In human hematopoietic cells, BMI-1 is preferentially expressed in primitive CD34+ bone marrow cells, whereas it is negative or very low in more mature CD34- cells (16). In peripheral lymphocytes, and particularly in follicular B cells, BMI-1 protein expression has been detected in resting cells of the mantle zone, whereas it is down-regulated in proliferating germinal center cells (17, 18). These observations indicate that BMI-1 expression in normal hematopoietic cells is tightly regulated in relation with cell differentiation in bone marrow and antigen-specific response in peripheral lymphocytes. BMI-1 expression in human tumors has not been examined previously. In this study, we have demonstrated that BMI-1 mRNA and protein expression show a distinct pattern in different types of lymphomas. Thus, BMI-1 levels were low in LCLs and FLs and significantly higher in MCLs and CLLs. These findings suggest that BMI-1 expression patterns in B cell lymphomas maintain in part the expression profile of their normal cell counterparts; because FLs and at least a subgroup of LCLs are considered lymphomas derived from follicular germinal center cells, whereas MCLs and CLLs are tumors mainly derived from naive pregerminal center cells. However, the four MCLs with BMI-1 gene amplification expressed significantly higher mRNA levels than all other tumors. In addition, five MCLs with no structural alterations of the gene showed high mRNA levels similar to those observed in tumors with BMI-1 gene amplification, suggesting that other mechanisms may be involved in up-regulation of the gene in these lymphomas. Different studies using animal models have shown a dose-dependent effect of BMI-1 gene expression on skeleton development

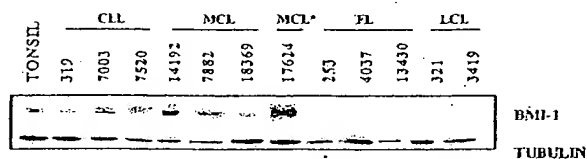


Fig. 3. Western blot analysis of BMI-1 protein in NHLs. The amplified MCL (17624) showed the highest BMI-1 protein levels, whereas other MCLs and CLLs had intermediate levels of expression. Very low or negative signal was observed in FLs and LCLs.

and lymphomagenesis (1, 3). These observations suggest that the high mRNA and protein levels detected in a subset of MCLs may play a role in the pathogenesis of these neoplasms.

Recent studies have identified the *INK4a/ARF* locus as a downstream target of the BMI-1 transcriptional repressor activity, suggesting that BMI-1 overexpression may contribute to human neoplasias that retain the wild-type *INK4a/ARF* locus (5). Interestingly, in our study, BMI-1 amplification and overexpression appeared in tumors with no alterations in *p16<sup>INK4a</sup>* and *p14<sup>ARF</sup>* genes. However, we could not detect differences in the expression levels of *p16<sup>INK4a</sup>* in tumors with and without BMI-1 gene alterations. The reasons for this apparent discrepancy with experimental observations are not clear. One possibility may be that genes other than *INK4a/ARF* are the main targets of BMI-1 repressor activity in these tumors. Particularly, different genes of the HOX family are regulated by BMI-1 and may also be involved in lymphomagenesis (19, 20).

In conclusion, the findings of this study indicate that BMI-1 gene expression is differentially regulated in B cell lymphomas. Alterations of the gene seem to be an uncommon phenomenon in human neoplasms, but they may contribute to the pathogenesis in a subset of MCLs. Although, BMI-1 gene alterations occurred in tumors with wild-type *INK4a/ARF* locus, the possible cooperation between these genes and the oncogenic mechanisms of BMI-1 in human neoplasms require additional analysis.

#### Acknowledgments

The authors thank Iracema Nayach for her excellent technical assistance.

#### References

- Haupt, Y., Alexander, W. S., Barri, G., Klinken, S. P., and Adams, J. M. Novel zinc finger gene implicated as myc collaborator by retrovirally accelerated lymphomagenesis in E  $\mu$ -myc transgenic mice. *Cell*, 65: 753-763, 1991.
- van Lohuizen, M., Verbeck, S., Scheijen, B., Wientjens, E., van der Gulden, H., and Berns, A. Identification of cooperating oncogenes in E  $\mu$ -myc transgenic mice by provirus tagging. *Cell*, 65: 737-752, 1991.
- Alkema, M. J., Jacobs, H., van Lohuizen, M., and Berns, A. Perturbation of B and T cell development and predisposition to lymphomagenesis in E $\mu$ Bmi1 transgenic mice require the Bmi1 RING finger. *Oncogene*, 15: 899-910, 1997.
- Haupt, Y., Bath, M. L., Harris, A. W., and Adams, J. M. Bmi-1 transgene induces lymphomas and collaborates with myc in tumorigenesis. *Oncogene*, 8: 3161-3164, 1993.
- Jacobs, J. J., Kieboom, K., Marino, S., DePinho, R. A., and van Lohuizen, M. The oncogene and Polycomb-group gene *bmi-1* regulates cell proliferation and senescence through the *ink4a* locus. *Nature (Lond)*, 397: 164-168, 1999.
- Alkema, M. J., Wiegant, J., Raap, A. K., Berns, A., and van Lohuizen, M. Characterization and chromosomal localization of the human proto-oncogene *Bmi-1*. *Hum. Mol. Genet.*, 2: 1597-1603, 1993.
- Pui, C. H., Raimondi, S. C., Murphy, S. B., Ribeiro, R. C., Kalwinsky, D. K., Dahl, G. V., Crist, W. M., and Williams, D. L. An analysis of leukemic cell chromosomal features in infants. *Blood*, 69: 1289-1293, 1987.
- Berger, R., Baranger, L., Bernheim, A., Valensi, F., Flandrin, G., and Berheim, A. T. Cytogenetics of T-cell malignant lymphoma. Report of 17 cases and review of the chromosomal breakpoints. *Cancer Genet. Cytogenet.*, 36: 123-130, 1988.
- D'Alessandro, E., Paterlini, P., Lo Re, M. L., Di Cola, M., Ligas, C., Quagliaro, D., and Del Porto, G. Cytogenetic follow-up in a case of Sezary syndrome. *Cancer Genet. Cytogenet.*, 45: 231-236, 1990.
- Bea, S., Ribas, M., Hernandez, J. M., Bosch, F., Pinyol, M., Hernandez, L., Garcia, J. L., Flores, T., Gonzalez, M., Lopez-Guillermo, A., Piris, M. A., Cardesa, A., Montserrat, E., Mito, R., and Campo, E. Increased number of chromosomal imbalances and high-level DNA amplifications in mantle cell lymphoma are associated with blastoid variants. *Blood*, 93: 4365-4374, 1999.
- Knuutila, S., Bjorkqvist, A. M., Autio, K., Tarkkanen, M., Wolf, M., Monni, O., Szymanska, J., Larramendi, M. L., Tapper, J., Pere, H., el-Rifai, W., Hemmer, S., Wasenius, V. M., Vidgren, V., and Zhu, Y. DNA copy number amplifications in human neoplasms: review of comparative genomic hybridization studies. *Am. J. Pathol.*, 152: 1107-1123, 1998.
- Alkema, M. J., Bronk, M., Verhoeven, E., Otte, A., van't Veer, L. J., Berns, A., and van Lohuizen, M. Identification of Bmi1-interacting proteins as constituents of a multimeric mammalian polycomb complex. *Genes Dev.*, 11: 226-240, 1997.
- Pinyol, M., Hernandez, L., Martinez, A., Cobo, F., Hernandez, S., Bea, S., Lopez-Guillermo, A., Nayach, I., Palacin, A., Nadal, A., Fernandez, P., Montserrat, E., Cardesa, A., and Campo, E. *INK4a/ARF* locus alterations in human non-Hodgkin's lymphomas mainly occur in tumors with wild type *p53* gene. *Am. J. Pathol.*, 156: 1987-1996, 2000.
- Pinyol, M., Cobo, F., Bea, S., Jares, P., Nayach, I., Fernandez, P. L., Montserrat, E., Cardesa, A., and Campo, E. *p16<sup>INK4a</sup>* gene inactivation by deletions, mutations, and hypermethylation is associated with transformed and aggressive variants of non-Hodgkin's lymphomas. *Blood*, 91: 2977-2984, 1998.
- Foot, A. B., Oakhill, A., and Kitchen, C. Acute monoclonal leukemia of infancy in Klinefelter's syndrome. *Cancer Genet. Cytogenet.*, 61: 99-100, 1992.
- Lessard, J., Baban, S., and Sauvageau, G. Stage-specific expression of polycomb group genes in human bone marrow cells. *Blood*, 91: 1216-1224, 1998.
- Raaphorst, F. M., van Kemenade, F. J., Fieret, E., Hamer, K. M., Salijn, D. P., Otte, A. P., and Meijer, C. J. Cutting edge: polycomb gene expression patterns reflect distinct B cell differentiation stages in human germinal centers. *J. Immunol.*, 164: 1-4, 2000.
- Raaphorst, F. M., van Kemenade, F. J., Blokzijl, T., Fieret, E., Hamer, K. M., Salijn, D. P., Otte, A. P., and Meijer, C. J. Coexpression of *Bmi-1* and *EZH2* polycomb group genes in Reed-Sternberg cells of Hodgkin's disease. *Am. J. Pathol.*, 157: 709-715, 2000.
- Gould, A. Functions of mammalian Polycomb group and trithorax group related genes. *Curr. Opin. Genet. Dev.*, 7: 488-494, 1997.
- van Oostveen, J., Bijl, J., Raaphorst, F., Walboomers, J., and Meijer, C. The role of homeobox genes in normal hematopoiesis and hematological malignancies. *Leukemia*, 13: 1675-1690, 1999.

# Gene-expression profiles predict survival of patients with lung adenocarcinoma

DAVID G. BEER<sup>1</sup>, SHARON L.R. Kardia<sup>2</sup>, CHIANG-CHING HUANG<sup>3</sup>, THOMAS J. GIORDANO<sup>4</sup>, ALBERT M. LEVIN<sup>2</sup>, DAVID E. MISEK<sup>5</sup>, LIN LIN<sup>1</sup>, GUOAN CHEN<sup>1</sup>, TAREK G. GHARIB<sup>1</sup>, DAFYDD G. THOMAS<sup>4</sup>, MICHELLE L. LIZYNESS<sup>4</sup>, RORK KUICK<sup>5</sup>, SATORU HAYASAKA<sup>1</sup>, JEREMY M.G. TAYLOR<sup>1</sup>, MARK D. IANNETTONI<sup>1</sup>, MARK B. ORRINGER<sup>1</sup> & SAMIR HANASH<sup>5</sup>

<sup>1</sup>Departments of <sup>1</sup>Surgery, <sup>2</sup>Epidemiology, <sup>3</sup>Biostatistics, <sup>4</sup>Pathology and <sup>5</sup>Pediatrics, University of Michigan, Ann Arbor, Michigan, USA

Correspondence should be addressed to D.G.B.; email: dgbeer@umich.edu.

Published online: 15 July 2002, doi:10.1038/nm733

Histopathology is insufficient to predict disease progression and clinical outcome in lung adenocarcinoma. Here we show that gene-expression profiles based on microarray analysis can be used to predict patient survival in early-stage lung adenocarcinomas. Genes most related to survival were identified with univariate Cox analysis. Using either two equivalent but independent training and testing sets, or 'leave-one-out' cross-validation analysis with all tumors, a risk index based on the top 50 genes identified low-risk and high-risk stage I lung adenocarcinomas, which differed significantly with respect to survival. This risk index was then validated using an independent sample of lung adenocarcinomas that predicted high- and low-risk groups. This index included genes not previously associated with survival. The identification of a set of genes that predict survival in early-stage lung adenocarcinoma allows delineation of a high-risk group that may benefit from adjuvant therapy.

Lung cancer remains the leading cause of cancer death in industrialized countries. Most patients with non-small cell lung cancer (NSCLC) present with advanced disease, and despite recent advances in multi-modality therapy, the overall 10-year survival rate remains a dismal 8–10%. However, a significant minority of patients (~25–30%) with NSCLC have stage I disease and receive surgical intervention alone. Although 35–50% of patients with stage I disease will relapse within 5 years<sup>1,2</sup>, it is not currently possible to identify specific high-risk patients.

Adenocarcinoma is currently the predominant histological subtype of NSCLC (refs. 1,5,6). Although morphological assessment of lung carcinomas can roughly stratify patients, there is a need to identify patients at high risk for recurrent or metastatic disease. Preoperative variables that affect survival of patients with NSCLC have been identified<sup>3–10</sup>. Tumor size, vascular invasion, poor differentiation, high tumor-proliferative index and several genetic alterations, including *K-ras* (refs. 11,12) and *p53* (refs. 10,13) mutations, have prognostic significance. Multiple independently assessed genes or gene products have also been investigated to better predict patient prognosis in lung cancer<sup>14–16</sup>. Technologies that simultaneously analyze the expression of thousands of genes<sup>17</sup> can be used to correlate gene-expression patterns with numerous clinical parameters—including patient outcome—to better predict tumor behavior in individual patients<sup>18</sup>. Analyses of lung cancers using array technologies have identified subgroups of tumors that differ according to tumor type and histological subclasses and, to a lesser extent, survival among adenocarcinoma patients<sup>19,20</sup>. Here we correlated gene-expression profiles with clinical outcome in a cohort of patients with lung adenocarcinoma and identified specific genes that

predict survival among patients with stage I disease. For further validation, we also show that the risk index predicted survival in an independent cohort of stage I lung adenocarcinomas.

## Hierarchical profile clustering yields three tumor subsets

Using oligonucleotide arrays, we generated gene-expression profiles for 86 primary lung adenocarcinomas, including 67 stage I and 19 stage III tumors, as well as 10 non-neoplastic lung samples. Selected sample replicates showed high correlation among coefficients and reliable reproducibility. We determined transcript abundance using a custom algorithm and the data set was trimmed of genes expressed at extremely low levels, that is, genes were excluded if the measure of their 75th percentile value was less than 100. Although potentially resulting in the loss of some information, trimming in this manner decreased the possibility that the clustering algorithm would be strongly influenced by genes with little or no expression in these samples. Hierarchical clustering with the resulting 4,966 genes yielded 3 clusters of tumors (Fig. 1). All 10 non-neoplastic samples clustered tightly together within Cluster 1 (data not shown). We examined the relationships between cluster and patient and tumor characteristics (Fig. 1 and Supplementary Figure A online). There were associations between cluster and stage ( $P = 0.030$ ) and between cluster and differentiation ( $P = 0.01$ ). Cluster 1 contained the greatest percentage (42.8%) of well differentiated tumors, followed by Cluster 2 (27%) and Cluster 3 (4.7%). Cluster 3 contained the highest percentage of both poorly differentiated (47.6%) and stage III tumors (42.8%), yet contained 3 (14.3%) moderately differentiated and 1 (5%) well differentiated stage I tumor. Notably, 11 stage I tumors were present in Cluster 3, sug-



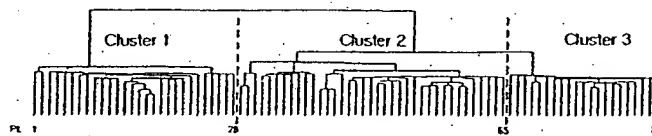
gesting a common gene-expression profile for this subset of stage I and stage III tumors.

For patients with stage I and stage III tumors, the average ages were 68.1 and 64.5 years and the percentage of smokers was 88.9% and 89.5%, respectively. Marginally significant associations between cluster and smoking history were observed ( $P = 0.06$ ). A significant relationship between histopathological classification and cluster was only discernable for bronchioloalveolar adenocarcinomas (BAs), which were only present in Clusters 1 and 2 ( $P = 0.0055$ ) and comprised 35.7% and 12.3% of tumors for Clusters 1 and 2, respectively.

We examined the heterogeneity in gene-expression profiles based on the trimmed data set among normal lung samples and stage I and stage III adenocarcinomas by calculating correlation coefficients between all pairs of samples. In contrast to normal lung samples that displayed highly similar gene-expression profiles (median correlation, 0.9), both stage I and III lung tumors demonstrated much greater heterogeneity in their expression profiles with lower correlation coefficients (median values, 0.82 and 0.79, respectively).

#### Northern-blot and immunohistochemistry analyses

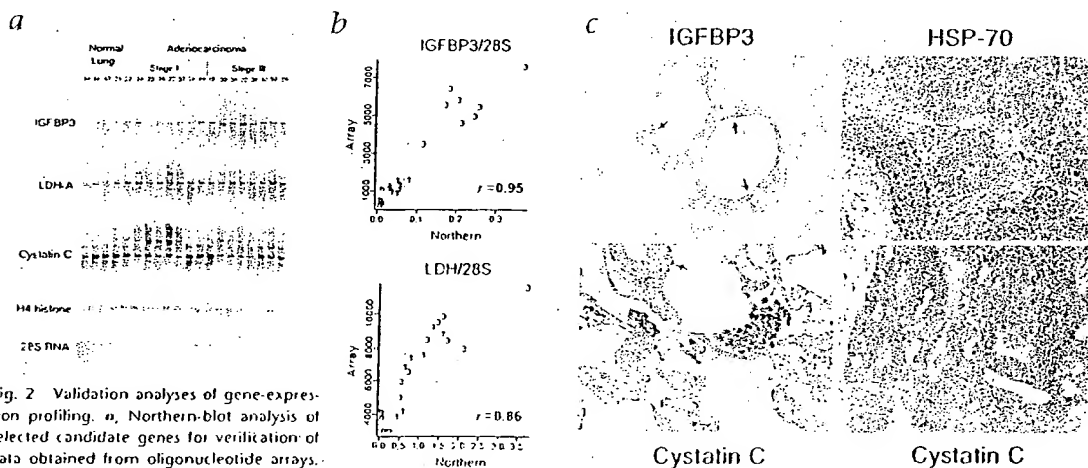
Of the 4,966 genes examined, 967 differed significantly between stage I and III adenocarcinomas, a number in excess of that expected by chance alone (248 at alpha level ( $\alpha$ ) = 0.05). Three genes were arbitrarily selected to verify the microarray expression data. The mRNA from 20 of the normal lung and tumor samples was examined by northern-blot hybridization with probes for insulin-like growth factor-binding protein 3 (IGFBP3), cystatin C



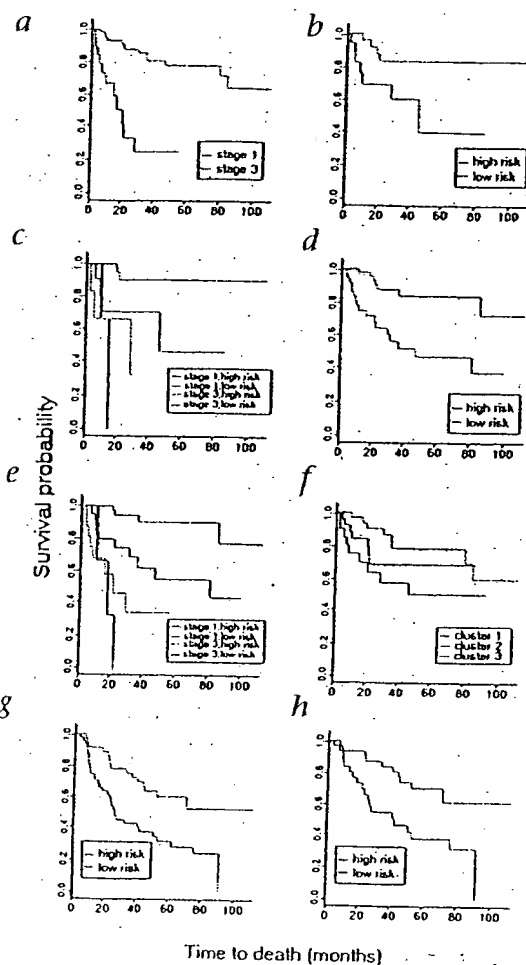
**Fig. 1** Unsupervised classification analysis of lung adenocarcinomas. 3 classes of tumors identified by agglomerative hierarchical clustering of gene-expression profiles using the 4,966 expressed genes. Patient and histopathological information for each lung adenocarcinoma case by cluster designation and methods for *K-ras* T2/T3th-codon mutational status and nuclear p53 protein accumulation are provided (Supplementary Figure A online). TN classification denotes information regarding patient tumor size and nodal involvement. Associations between cluster membership and patient or histopathological variables are indicated at significance level ( $P \leq 0.05$ ).

and lactate dehydrogenase A (LDH-A) (Fig. 2a). Two gene probes not represented on the microarrays were used as controls, including histone H4, a potential index of overall cell proliferation, and 28S ribosomal RNA, a control for sample loading and transfer. The relative amounts of IGFBP3, cystatin C and LDH-A mRNA strongly correlated with microarray-based measurements (Fig. 2b). In both assays, IGFBP3 and LDH-A mRNA levels increased from stage I to stage III adenocarcinomas and were higher than those in normal lung. Cystatin C mRNA levels were more variable but relatively greater in normal lung than tumors. These results suggest that the oligonucleotide microarrays provided reliable measures of gene expression. The tumors showed slightly greater histone H4 expression than the normal lung, likely reflecting increased proliferation of tumor cells.

Immunohistochemistry was performed for IGFBP3, cystatin C and HSP-70 to determine whether mRNA overexpression was reflected by an increase of their corresponding proteins in tumors.



**Fig. 2** Validation analyses of gene-expression profiling. **a**, Northern-blot analysis of selected candidate genes for verification of data obtained from oligonucleotide arrays. The same sample RNA for the 4 uninvolved lung, 8 stage I and 8 stage III tumors was used for the northern-blot and oligonucleotide array analyses. **b**, Correlation analysis of quantitative data obtained from oligonucleotide arrays and northern blots measured by integrated phosphorimager based signals for the IGFBP3 and LDH-A genes. The ratio of IGFBP3, cystatin C and LDH-A mRNA to 28S rRNA was determined. The relative values for each gene from each sample are shown, n, non-neoplastic normal lung; 1, stage I tumors; 3, stage III tumors. **c**, Immunohistochemical analysis of IGFBP3, HSP-70 and cystatin C in lung and lung adenocarcinomas. Cytoplasmic IGFBP3 immunoreactivity in a neoplastic gland (tumor 127) with prominent apical staining (blue reactant staining, arrow, upper left). Diffuse cytoplasmic HSP-70 immunoreactivity (tumor 127), yet stromal elements show no reactivity (upper right). Normal lung parenchyma (lower left) shows cytoplasmic cystatin C immunoreactivity in alveolar pneumocytes (arrow) and intra-alveolar macrophages but tumor (190) shows diffuse cytoplasmic cystatin C immunoreactivity with prominent apical staining (lower right). Magnification,  $\times 200$ .



**Fig. 3** Gene-expression profiles and patient survival. **a**, Relationship between tumor stage and patient survival (stage I and stage III differ significantly,  $P < 0.0001$ ). **b**, Relationship between the survival in the 43 test samples and their risk assignments based on the 50-gene risk index estimated in the 43 training samples. The high- and low-risk groups differ significantly ( $P = 0.024$ ). **c**, Relationship between patient survival and the risk assignments in test samples (in **b**) conditional for tumor stage. The high- and low-risk stage I groups differ significantly ( $P = 0.028$ ), whereas stage III low- and high-risk groups did not ( $P = 0.634$ ). **d**, Relationship between survival in the test cases and their risk assignments based on the 86 'leave-one-out' cross-validation of the 50-gene risk index. The high- and low-risk groups differ significantly ( $P = 0.0006$ ). **e**, Relationship between test case's risk assignment and survival (in **d**) conditional on tumor stage. The high- and low-risk stage I lung adenocarcinoma groups differ significantly from each other ( $P = 0.003$ ), whereas low- and high-risk stage III tumors do not. **f**, Relationship between tumor class identified by hierarchical clustering and patient survival. Survival for patients in Cluster 3 differed relative to the tumors in Cluster 2 ( $P = 0.037$ ) and approached significance for Cluster 1 and 2 combined ( $P = 0.06$ ). **g**, Analysis of the Michigan-based risk index using top cross-validated survival genes identify a low- and high-risk group in an independent cohort of 84 Massachusetts-based lung adenocarcinomas that are significantly different ( $P = 0.003$ ). **h**, Among the 62 stage I lung adenocarcinomas in the Massachusetts sample, the high- and low-risk groups differed significantly ( $P = 0.006$ ).

After conservatively choosing the 60th percentile cutoff point from the training set, we then applied this risk index and cutoff point to the testing set. The risk index of the top 50 genes correctly identified low- and high-risk individuals within the independent testing set ( $P = 0.024$ ) (Fig. 3b and Supplementary Methods online). Notably, 11 stage I tumors were included in the high-risk subgroup. When this risk assignment was then conditionally examined for stage progression (Fig. 3c), low- and high-risk groups among stage I tumors were found to differ ( $P = 0.028$ ) in their survival.

#### Identification of a robust set of survival genes

Although predictive of patient survival, a single training-testing set may not provide the most robust set of genes due to random sampling issues. Therefore, a 'leave-one-out' cross-validation approach was used to identify genes associated with survival from all 86-tumor samples. We first developed a 50-gene risk index in each training set, and then applied the risk index to the test case held out from the full set of tumors and assigned the held out tumor to the high- or low-risk groups (Fig. 3d). The high- and low-risk subgroups determined in the test cases differed significantly in their overall survival ( $P = 0.0006$ ). Among the larger group of stage I lung adenocarcinomas, the low-risk ( $n = 46$ ) and high-risk ( $n = 21$ ) groups had markedly different survival ( $P = 0.003$ ) (Fig. 3e). Table 1 lists selected examples of the cumulative top 100 genes derived from this cross-validation procedure (complete list in Supplementary Table A online).

It was also noted that many of the stage I patients in the high-risk subgroup (Fig. 3e) were present in Cluster 3 (Fig. 1). Kaplan-Meier analysis (Fig. 3f) demonstrated a significantly worse survival ( $P = 0.037$ ) for patients in Cluster 3 relative to patients in Cluster 2 and approaching significance for Cluster 1 and 2 combined ( $P = 0.06$ ). This further indicates the important relationship between gene-expression profiles and patient survival, independent of disease stage.

Consistent with previous analyses of lung adenocarcinomas<sup>23</sup>, 40% of stage I and 57.8% of stage III tumors had 12th or 43th codon *K-ras* gene mutations. Those patients with tumors containing *K-ras* mutations showed a trend of poorer survival, but

Immunoreactivity for both *IGFBP-3* and *HSP-70* (Fig. 2c) was detected in the cytoplasm of the adenocarcinomas, with little detectable reactivity in the stromal or inflammatory cells. Cystatin C was detected in alveolar pneumocytes and intra-alveolar macrophages in non-neoplastic lung parenchyma and also consistently in the cytoplasm of neoplastic cells.

#### Gene-expression profiles predict survival

As expected, Kaplan-Meier survival curves (Fig. 3a) and log-rank tests indicated poorer survival among stage III compared with stage I adenocarcinomas ( $P = < 0.0001$ ). Two statistical approaches were used to determine whether gene-expression profiles could predict survival using the data set of 4,966 genes. In one approach, equal numbers of randomly assigned stage I and stage III tumors constituted training ( $n = 43$ ) and testing ( $n = 43$ ) sets. In the training set, the top 10, 20, 50 or 75 genes were used to create risk indices that were evaluated for their association with survival using the 50th, 60th or 70th percentile cutoff points to categorize patients into high or low groups. The results were similar across cutoff points but the 50-gene risk index had the best overall association with survival in the training set.





Table 1 Selected examples of the top 100 genes from cross-validation

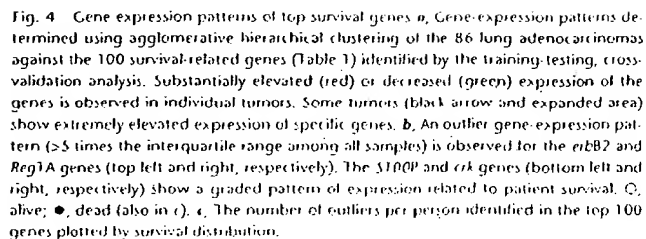
Gene name	P (normal versus tumor t-test)	% Change in tumor	P (stage I versus stage III t-test)	% Change in stage III	Coefficient $\beta$	Unigene comment
CASP4	0.56	-6%	0.02	57%	0.0022	Apoptosis-related Caspase 4, apoptosis- related cysteine protease
P63	9.73E-04	37%	0.03	43%	0.0010	Transmembrane protein (63 kD), endoplasmic reticulum/ Golgi intermediate compartment
KRT7	8.02E-08	126%	0.11	55%	0.0003	Cell adhesion and structure Keratin 7
LAMB1	0.14	-20%	0.01	60%	0.0027	Laminin, $\beta$ 1
BMP2	0.54	-21%	0.27	47%	0.0044	Cell cycle and growth regulators Bone morphogenetic protein 2
CDC6	1.31E-05	1070%	0.05	148%	0.0124	CDC6 (cell division cycle 6, <i>Saccharomyces cerevisiae</i> homolog)
S100P	2.10E-08	1572%	0.19	77%	0.0001	S100 calcium-binding protein P
SERPINE1	2.89E-03	72%	0.25	30%	0.0008	Serine (or cysteine) proteinase inhibitor, clade E (nexin)
STX1A	8.65E-08	54%	0.07	26%	0.0031	Syntaxin 1A (brain)
ADM	0.05	39%	0.04	117%	0.0016	Cell signaling adrenomedullin
AKAP 12	8.53E-03	-47%	0.05	214%	0.0010	A kinase (PRKA) anchor protein (gravin) 12
ARHGE	0.06	-39%	0.05	87%	0.0092	ras homolog gene family, member E
GRB7	2.02E-03	38%	0.63	15%	0.0030	Growth factor receptor-bound protein 7
VEGF	6.50E-08	174%	0.02	85%	0.0013	Vascular endothelial growth factor
WNT10B	0.05	31%	0.48	20%	0.0022	Wingless-type MMTV integration site family, member 10B
HSPA8	0.36	8%	9.01E-04	51%	0.0008	Chaperones Heat-shock 70 kD protein 8
ERBB2	0.04	92%	0.37	120%	0.0013	Receptors v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 2
FXR3	0.10	111%	0.31	73%	0.0046	FXR domain-containing ion transport regulator 3
SLC20A1	1.34E-03	58%	0.02	66%	0.0021	Solute carrier family 20 (phosphate transporter), member 1
CSTB	1.57E-04	50%	0.15	34%	0.0001	Enzymes, cellular metabolism Cystatin B (stelin B)
CTSL	0.48	-10%	0.03	67%	0.0007	Cathepsin L
CYP24	3.16E-06	N/A	0.97	2%	0.0008	Cytochrome P450, subfamily XXIV (vitamin D 24-hydroxylase)
FUT3	1.07E-07	114%	0.97	-1%	0.0033	Fucosyltransferase 3 (galactoside 3(4)-L- fucosyltransferase, Lewis blood group included)
MLN64	0.20	32%	0.42	80%	0.0007	Steroidogenic acute regulatory protein related
PDE7A	0.12	33%	0.01	-35%	-0.0187	Phosphodiesterase 7A
PLGL	0.04	-68%	0.35	-170%	-0.0011	Plasminogen-like
SLC1A6	0.07	-32%	0.12	86%	0.0069	Solute carrier family 1 (high-affinity aspartate/ glutamate transporter), member 6
COPIB	0.10	-33%	0.26	25%	0.0016	Transcription and translation Core promoter element binding protein
CRK	0.10	32%	0.03	48%	0.0098	v-crk avian sarcoma virus CT10 oncogene homolog
RELA	0.26	-7%	0.01	20%	0.0034	v-rel avian reticuloendotheliosis viral oncogene homolog A
KIAA0005	2.71E-04	40%	0.02	45%	0.0010	Unknown function KIAA0005 gene product
MCB1	0.27	125%	0.33	459%	0.0018	Mammaglobin 1

Bolded genes were also significant for survival in 43 tumor training set (Fig. 3b).

Table 1 Selected examples of the cumulative top 100 genes identified using training-testing, cross-validation of all 86 lung tumor samples. The percent change, as well as the direction, for the average values of the 10 non-neoplastic lung to all tumors, and for the 67 stage I to the 19 stage III tumors are shown. A positive coefficient  $\beta$  value is indicative of a relationship of gene expression to a

poorer patient outcome. The genes are listed in potential functional categories. Genes that were also present in the top 50 survival genes using the 43-tumor training set (Fig. 3b) are indicated in bold type. Complete listing of the gene probe sets and annotated gene and unigene identifiers can be found in the Supplementary Methods.





this difference did not reach statistical significance among all patients ( $P = 0.25$ ), between patients within tumor clusters ( $P = 0.41$ ) or when analyzed separately among stage I ( $P = 0.22$ ) and stage III ( $P = 0.53$ ) patients. Nuclear accumulation of p53 was detected in 17.9% stage I and in 22.2% stage III tumors. No significant relationship was observed for p53 staining and patient survival, cluster or tumor stage.

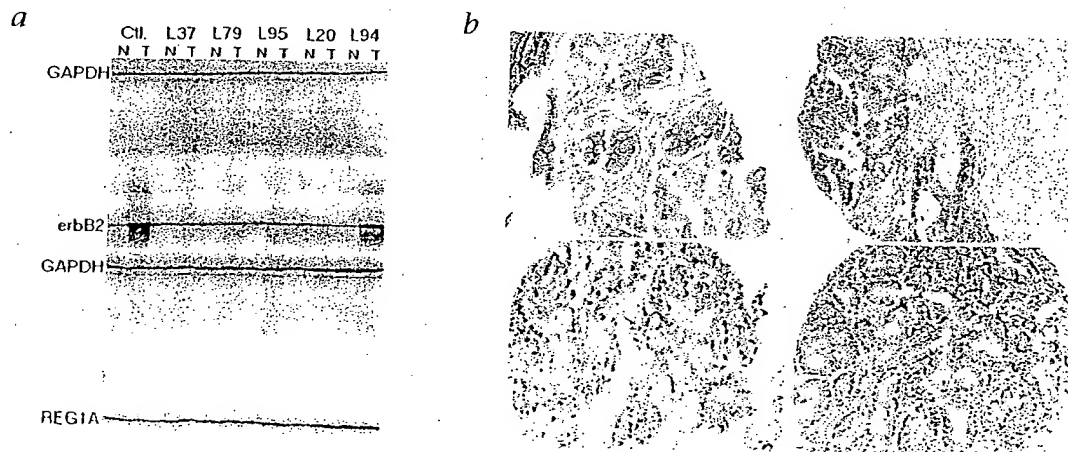
**Confirmation using an independent set of adenocarcinomas**  
The robustness of our 50-gene risk index in predicting survival in lung adenocarcinomas was tested using oligonucleotide gene-expression data obtained from a completely independent (Massachusetts-based) sample of 84 lung adenocarcinomas (62 stage I, 14 stage II and 8 stage III; ref. 21, and dataset A at [www.genome.wi.mit.edu/MPR/lung](http://www.genome.wi.mit.edu/MPR/lung)). To ensure equivalent power for testing and comparability of samples, the criteria for including tumors in the analysis were 40% or greater tumor cellularity, no mixed histology (that is, adenosquamous) and patient survival information. To obtain comparative gene-expression measures between the two data sets, gene sequences present on the U95A and HuGeneFL array were examined, and expression data for our top 50 cross-validation genes for all 84 Massachusetts samples were obtained and processed<sup>24</sup> (see also Supplementary Methods online). When we examined the risk assignment of these 84 samples, employing the identical cutoff point used for the 86 Michigan-based lung samples, we observed low- and high-risk groups (Fig. 3g;  $P = 0.003$ ). Notably, among the 62 stage I tumors, high- and low-risk groups were observed that differed significantly ( $P = 0.006$ ) in their survival (Fig. 3h).

**Survival genes had graded and outlier expression patterns**  
A statistical and graphical analysis of the 100 survival-related

genes (Table 1) clustered against all 86 tumors revealed individual tumors with substantially elevated expression in both a limited and larger number of genes (Fig. 4a). Among these genes, we observed two distinct patterns of expression related to patient survival. One pattern, designated 'outlier', included genes showing substantially elevated expression (greater than five times the interquartile range among all samples), whereas the other pattern, designated 'graded', was characterized by continuously distributed expression with patient survival (Fig. 4b). The *erbB2* and *Reg1A* genes are examples of outlier expression patterns and *S100P* and *crk* genes of graded patterns. The number of outliers per person in the top 100 genes was identified and plotted according to survival times and events (Fig. 4c). Both stage I and stage III lung adenocarcinomas showed outlier gene patterns and 10 tumors contained 3 or more outlier genes.

Because gene amplification may result in increased gene expression, the nine genes with outlier expression patterns (*erbB2*, *SLC1A6*, *Wnt 1*, *MGB1*, *Reg1A*, *AKAP12*, *PACE*, *CYP24*, *KYNU*) and one gene with a graded expression pattern (*KRT18*) were examined using quantitative genomic PCR to evaluate genomic copy number (Fig. 5a). Gene amplification of *erbB2* (17q12) was detected in tumor L94, which had the highest *erbB2* mRNA expression (Fig. 4a). Gene amplification was not detected for any of the other seven tested genes in tumor L94, as well as in other tumors. The two genes most frequently demonstrating the outlier pattern in these lung adenocarcinomas were *KYNU* and *CYP24*, and were present in 10 and 9 tumors, respectively. *CYP24* has been described as a gene amplified and overexpressed in breast cancer<sup>25</sup>, and these results indicate elevated expression in lung adenocarcinoma.

To determine whether the graded or outlier gene-expression patterns also occur at the protein-expression level, 10 of the 100



**Fig. 5** Gene amplification and protein expression of survival-related genes. **a**, Analysis of potential gene amplification for 9 genes showing outlier expression patterns in the lung tumors (*erbB2*, *SLC1A6*, *Wnt 1*, *MGB1*, *Reg1A*, *AKAP12*, *PACE*, *CYP24* and *KYNU*) and examined using quantitative genomic PCR. A gene showing graded expression pattern (*KRT18*), and one gene (*PACE4*) with a similar chromosome location as *PACE1*, were used as controls. Only *erbB2* and *Reg1A* are shown. An esophageal adenocarcinoma with known high-level genomic amplification of *erbB2* was used as a positive control and normal esophagus DNA was used as a negative control (C11). PCR fragments sizes were 343 bp for GAPDH, 166 bp for *erbB2* and 126 bp for

*Reg1A*. DNA is from normal lung (N) and tumor (T) from each patient (for example L37). **b**, Immunohistochemical analysis of survival-related genes with lung adenocarcinoma microarrays using the tumors from this study. The transmembrane *erbB2* protein (top left) expression is substantially increased in tumor L94 containing the amplified *erbB2* gene (Fig. 4a and b). Expression of VEGF (top right) and S100P (bottom left) was located within the neoplastic cells and the pattern of immunoreactivity was consistent with the graded expression pattern demonstrated by their mRNA profiles. Expression of the oncogene *crk* (bottom right) was abundantly expressed in neoplastic lung cells. Magnification,  $\times 400$  (*erbB2*);  $\times 200$  (VEGF, S100P and *crk*).

## ARTICLES

top survival genes (Table 1) for which specific antibodies were available were chosen for immunohistochemical analysis using lung-tumor arrays from this study (Fig. 5b). Expression of membrane *erbB2* protein was substantially increased in the *erbB2*-amplified tumor L94 and very low levels of expression were present in other tumors, consistent with mRNA-expression measurements (Fig. 4a and b). CDC6 protein expression was also substantially higher in tumor L94, consistent with mRNA levels (data not shown). Expression of vascular endothelial growth factor (VEGF) and S100P (Fig. 5b), as well as cytokeratin 18 (KRT18), cytokeratin 7 (KRT7) and *fas*-associated death domain (FADD) protein (data not shown), was located within the lung tumor cells and consistent with the graded expression pattern of the mRNA profiles. The oncogene *crk* showed both graded mRNA as well as a graded protein-expression pattern with survival, and was abundantly expressed in the tumor cells (Fig. 5b). These results indicate that many survival-associated genes are expressed at the protein level and demonstrate similar mRNA and protein-expression patterns.

### Discussion

We used several approaches for the analysis of gene-expression data related to clinicopathological variables and patient survival. One approach, hierarchical clustering, was used to examine similarities among lung adenocarcinomas in their patterns of gene expression. Previous studies of lung tumors<sup>21,22</sup> have also used this method to describe subclasses of lung tumors. Here, we found three clusters that showed significant differences with respect to tumor stage and tumor differentiation. This suggests, as expected, that tumors with similar histological features of differentiation demonstrate similarities in gene expression. This feature also partly underlies the observed statistical association of tumor stage and cluster, as many of the higher-stage tumors, often poorly differentiated and previously associated with a reduced survival<sup>9,10</sup>, were located in Cluster 3. Although this cluster contained the highest percentage of stage III tumors, it also contained a nearly equal mixture of stage I and stage III tumors and not all tumors were poorly differentiated. This indicates that a subset of stage I lung adenocarcinomas share gene-expression profiles with higher-stage tumors. Notably, 10 of the 11 stage I tumors found in Cluster 3 were the high-risk stage I tumors identified using the risk index in the 'leave-one-out' cross-validation.

In contrast to previous analyses of lung adenocarcinomas<sup>21,22</sup>, we validated the expression data from the arrays. The strong correlation of northern-blot analysis and oligonucleotide-array data for gene expression in the same samples (Fig. 2b) indicates that these studies provide robust gene-expression estimates. Immunohistochemistry using the same tumor samples in tissue arrays demonstrates protein expression within the lung tumor cells. Together, these studies indicate that many of the genes identified using gene-expression profiles are likely relevant to lung adenocarcinoma. For example, *IGFBP3* gene expression is increased in lung adenocarcinomas (Fig. 2c). *IGFBP3* protein modulates the autocrine or paracrine effects of insulin-like growth factors, elevated *IGFBP3* expression is observed in colon cancer<sup>24</sup>, and increased serum *IGFBP3* is associated with progression in breast cancer<sup>25</sup>. Heat-shock protein 70 (HSP-70) is increased in lung adenocarcinomas of smokers<sup>26</sup> and is associated with increased metastatic potential in breast cancer<sup>27</sup>. Increased serum lactate dehydrogenase is correlated with tumor stage and tumor burden<sup>28</sup>, and cystatin C, a cysteine protease inhibitor ex-

pressed in human lung cancers<sup>29</sup>, is prognostic in some cancers<sup>32</sup>. The decreased expression of this protease inhibitor may affect the invasive properties of the tumor cell.

The cross-validation analytical strategy we used is particularly informative for these types of gene-expression analyses for disease outcome<sup>33,34</sup>, and identification of cross-validated genes with a larger tumor cohort may help refine this risk index for use in a clinical setting. The gene-expression data also provide opportunities to observe overarching patterns that advance our understanding of associations between genes and disease. For example, the top 100 survival genes include those involved in signaling, cell cycle and growth, transcription, translation and metabolism. Expression of many of these genes is likely a function of increased proliferation and metabolism in the more aggressive tumors. Some genes, such as *erbB2* and *Reg1A* (Fig. 4a and b), were highly overexpressed in a few patients having poor survival. In one tumor, the *erbB2* gene was amplified (Fig. 5a), demonstrating that genomic changes may underlie the overexpression of a subset of these outlier genes. Immunohistochemistry confirmed protein overexpression in this patient's tumor (Fig. 5b). Notably, seven of the eight outlier genes were not amplified, indicating that other mechanisms underlie the increased mRNA expression of these survival-related genes.

Most genes showed a graded relationship between expression and patient survival. Genes such as that encoding VEGF, known to be strongly associated with survival in lung cancer<sup>35,36</sup>, were identified as related to patient survival in our study. VEGF demonstrated a graded expression pattern, as did the S100P and *crk* oncogene (Fig. 5b). S100P is a calcium-regulated protein not previously reported in lung cancer. The *crk* gene, the cellular homolog of the *v-crak* oncogene, is a member of a family of adaptor proteins involved in signal transduction and interacts directly with c-jun N-terminal kinase 1 (JNK1)<sup>37</sup>. Although *crk* has not been shown to have a role in lung cancer, its role in the MAP-kinase pathway, which leads to activation of matrix metalloproteinase secretion and cell invasion<sup>38</sup>, indicates potential involvement in the tumor cell invasion or metastasis of some lung adenocarcinomas. Among the many genes identified in this study, like *crk*, that may be causally involved in lung cancer progression (Table 1), some were related to survival in many patients, and others in only smaller subsets of patients. This result is consistent with the complex molecular architecture of tumors in general, the heterogeneity of lung adenocarcinomas in particular and the multiple mechanisms underlying tumor-cell survival, invasion and metastasis<sup>39</sup>.

Our results demonstrate that a gene-expression risk profile—based on the genes most associated with patient survival—can distinguish stage I lung adenocarcinomas and differentiate prognoses. The particular genes that define the clusters, or are associated with survival, likely reflect the characteristics of the particular tumors included in the analysis. Current therapy for patients with stage I disease usually consists of surgical resection without adjuvant treatment<sup>40</sup>. Clearly, the identification of a high-risk group among patients with stage I disease would lead to consideration of additional therapeutic intervention for this group, possibly leading to improved survival of these patients.

### Methods

**Patient population.** Sequential patients seen at the University of Michigan Hospital between May 1994 and July 2000 for stage I or stage III lung adenocarcinoma were evaluated for this study. Consent was received and the project was approved by the local Institutional Review Board. Primary tumors and adjacent non-neoplastic lung tissue were obtained at the time of



surgery. Peripheral portions of resected lung carcinomas were sectioned, evaluated by a study pathologist and compared with routine H&E sections of the same tumors, and utilized for mRNA isolation. Regions chosen for analysis contained a tumor cellularity greater than 70%, no mixed histology, potential metastatic origin, extensive lymphocytic infiltration or fibrosis. Tumors were histopathologically divided into two categories based on their growth pattern: bronchial-derived, if they exhibited invasive features with architectural destruction, and bronchioloalveolar, if they exhibited preservation of the lung architecture. All stage I patients received only surgical resection with intra-thoracic nodal sampling and no other treatments. Stage III patients received surgical resection plus chemotherapy and radiotherapy.

**Gene-expression profiling and K-ras mutation analysis.** RNA isolation, cDNA synthesis and gene-expression profiling were performed as described<sup>24</sup>. Details of gene annotation and K-ras mutation analysis are provided in supplementary information.

**Northern-blot analysis.** Total cellular RNA (10 µg) was separated in 1.2% agarose-formaldehyde gels and vacuum-transferred to Gene Screen Plus (NEN Life Science Products, Boston, Massachusetts). Hybridization conditions and probe labeling were as described<sup>25</sup>. Individual sequence-validated cDNA image clones for human *IGFBP3* (clone 1407750), *LDH-A* (clone 2420241), *cystatin C* (CT53; clone 949938) were from Research Genetics (Huntsville, Alabama). The human histone H4 cDNA and the 28S ribosomal RNA 26-mer oligonucleotide probe were prepared and labeled as described<sup>26</sup>.

**Gene-amplification analysis.** 11 genes were selected for the analysis of genomic alterations. Primers were designed using PrimerSelect 4.05 Windows 32 software (DNASTAR, Madison, Wisconsin), avoiding pseudogenes or potential homologous regions. Forward and reverse primers for the genes are provided (Supplementary Methods online). Quantitative genomic-PCR was then applied and analyzed as described<sup>27</sup>.

**Immunohistochemical staining.** The H&E-stained slides of all primary lung tumors were used to identify the most representative regions of each tumor and a tissue microarray (TMA) block was constructed as described<sup>28</sup>. Immunohistochemistry (IHC) was performed using both routine and sections from the TMA block as described<sup>29</sup>. Detailed methods and the concentrations used for all antibodies are provided in the Supplementary Methods.

**Statistical methods.** *t*-tests were used to identify differences in mean gene-expression levels between comparison groups. Agglomerative hierarchical clustering<sup>30</sup> was applied using the average linkage method to investigate whether there was evidence for natural groupings of tumor samples based on correlations between gene-expression profiles. To investigate the robustness of the clustering inference, gene-expression values were perturbed by adding random Gaussian error of magnitude obtained from a duplicate sample to each data point and then reclustered to determine concordance in the tumor's class membership. Pearson,  $\chi^2$  and Fisher's exact tests were used to assess whether cluster membership was associated with physical and genetic characteristics of the tumors.

To determine whether gene-expression profiles were associated with variability in survival times, 2 separate but complementary approaches were used. In the first approach, the 86 tumors were randomly assigned to equivalent training and testing sets consisting of equal numbers of stage I and III tumors in order to validate a novel risk-index function that captured the effect of many genes at once. In the second approach, cross-validation<sup>31</sup> was used to more robustly identify the genes associated with survival. Briefly, a 'leave-one-out' cross-validation procedure in which 85 of the 86 tumors (the training set) was used to identify genes that were univariately associated with survival. The risk index was defined as a linear combination of the gene-expression values for the top genes identified by univariate Cox proportional-hazard regression modeling<sup>32</sup>, weighted by their estimated regression coefficients. Kaplan-Meier survival plots and log-rank tests were then used to assess whether the risk-index assignment to high/low categories was validated in the test set. A more detailed description is provided (Supplementary Methods online).

Note: Supplementary information is available on the Nature Medicine website.

#### Acknowledgments

We thank D. Sonders for technical assistance; D. Sing for assistance with the figures; and G. Omenn for critical reading of this manuscript. This work was supported by National Cancer Institute grant: U19 CA-85953 and the Tissue Core of the University of Michigan Comprehensive Cancer Center (NIH CA-46952).

#### Competing interests statement

The authors declare that they have no competing financial interests.

RECEIVED 5 APRIL; ACCEPTED 14 JUNE 2002

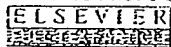
1. Fry, W.A.; Phillips, J.L. & Menck, H.R. Ten-year survey of lung cancer treatments and survival in hospitals in the United States. *Cancer* 66, 1867-1876 (1999).
2. Williams, D.E. et al. Survival of patients surgically treated for stage I lung cancer. *J. Thorac. Cardiovasc. Surg.* 82, 70-76 (1981).
3. Pairolero, P.C. et al. Posturgical stage I bronchogenic carcinoma: Morbid implications of recurrent disease. *Ann. Thorac. Surg.* 38, 331-338 (1984).
4. Naruke, T. et al. Prognosis and survival in resected carcinoma based on the new international staging system. *J. Thorac. Cardiovasc. Surg.* 96, 440-447 (1988).
5. Kaisermann, M.C. et al. Evolving features of lung adenocarcinoma in Rio de Janeiro, Brazil. *Oncol. Rep.* 8, 189-192 (2001).
6. Roggli, V.L. et al. Lung cancer heterogeneity: A blinded and randomized study of 100 consecutive cases. *Hum. Pathol.* 16, 569-579 (1985).
7. Gail, M.H. et al. Prognostic factors in patients with resected stage I non-small cell lung cancer: A report from the Lung Cancer Study Group. *Cancer* 54, 1802-1813 (1984).
8. Takise, A. et al. Histopathologic prognostic factors in adenocarcinomas of the peripheral lung less than 2 cm in diameter. *Cancer* 61, 2083-2088 (1988).
9. Ichinose, Y. et al. Is T factor of the TMN staging system a predominant prognostic factor in pathologic stage I non-small cell lung cancer. *J. Thorac. Cardiovasc. Surg.* 106, 90-94 (1993).
10. Harpole, D.H. et al. A prognostic model of recurrence and death in stage I non-small cell lung cancer utilizing presentation, histopathology, and oncoprotein expression. *Cancer Res.* 55, 51-56 (1995).
11. Rodenhuis, S. et al. Mutational activation of the K-ras oncogene: A possible pathogenic factor in adenocarcinoma of the lung. *N. Engl. J. Med.* 317, 929-935 (1987).
12. Slebos, R.J.C. et al. K-ras oncogene activation as a prognostic marker in adenocarcinoma of the lung. *N. Engl. J. Med.* 323, 561-565 (1990).
13. Horio, Y. et al. Prognostic significance of p53 mutations and 3p deletions in primary resected non-small cell lung cancer. *Cancer Res.* 53, 2-4 (1993).
14. Kern, J.A. et al. C-erbB-2 expression and codon 12 K-ras mutations both predict shortened survival for patients with pulmonary adenocarcinomas. *J. Clin. Invest.* 93, 516-520 (1994).
15. Ebina, M. et al. Relationship of p53 overexpression and up-regulation of proliferating cell nuclear antigen with the clinical course of non-small cell lung cancer. *Cancer Res.* 54, 2496-2503 (1994).
16. Mehdi, S.A. et al. Prognostic markers in resected stage I and II non-small cell lung cancer: an analysis of 260 patients with 5 year follow-up. *Clin. Lung Cancer* 1, 59-67 (1997).
17. Schneider, P.M. et al. Multiple molecular marker testing (p53, c-Ki-ras, c-erbB-2) improves estimation of prognosis in potentially curative resected non-small cell lung cancer. *Br. J. Cancer* 83, 473-479 (2000).
18. Herbst, R.S. et al. Differential expression of E-cadherin and type IV collagenase genes predicts outcome in patients with stage I non-small cell lung carcinoma. *Clin. Can. Res.* 6, 790-797 (2000).
19. Liotta, L. & Petricoin, E. Molecular profiling of human cancer. *Nature Rev. Genet.* 1, 48-56 (2000).
20. Golub, T.R. Editorial: Genome-wide views of cancer. *N. Engl. J. Med.* 344, 601-602 (2001).
21. Bhattacharjee, A. et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc. Natl. Acad. Sci. USA* 98, 13790-13795 (2001).
22. Garber, M.E. et al. Diversity of gene expression in adenocarcinoma of the lung. *Proc. Natl. Acad. Sci. USA* 98, 13784-13789 (2001).
23. Mills, N.L. et al. Increased prevalence of K-ras oncogene mutations in lung adenocarcinoma. *Cancer Res.* 55, 1444-1447 (1995).
24. Giordano, T.J. et al. Organ-specific molecular classification of lung, colon and ovarian adenocarcinomas using gene expression profiles. *Am. J. Pathol.* 159, 1231-1238 (2001).
25. Albertson, D.G. et al. Quantitative mapping of amplicon structure by array CGH identifies CYP24 as a candidate oncogene. *Nature Genet.* 25, 144-146 (2000).
26. Kanara, S. et al. IGFBP-3 mediates TGF- $\beta$ 1 proliferative response in colon cancer cells. *Int. J. Cancer* 87, 373-378 (2000).
27. Vadgama, J.V. et al. Plasma insulin-like growth factor-I and serum IGF-binding protein 3 can be associated with the progression of breast cancer, and predict the risk of recurrence and the probability of survival in African-American and Hispanic



## ARTICLES

- women. *Oncology* 57, 330–340 (1999).
28. Volm, M., Mattern, J. & Stämmler, G. Up-regulation of heat shock protein 70 in adenocarcinoma of the lung in smokers. *Anticancer Res.* 15, 2607–2609 (1995).
29. Ciocca, D.R. *et al.* Heat shock protein hsp70 in patients with axillary lymph node-positive breast cancer: prognostic implications. *J. Natl. Cancer Inst.* 85, 570–574 (1993).
30. Rotenberg, Z. *et al.* Total lactate dehydrogenase and its isoenzymes in serum of patients with non-small cell lung cancer. *Clin. Chem.* 34, 668–670 (1988).
31. Krepela, E. *et al.* Cysteine proteases and cysteine protease inhibitors in non-small cell lung cancer. *Neoplasma* 45, 318–331 (1998).
32. Kos, J. *et al.* Cysteine proteinases and their inhibitors in extracellular fluids: Markers for diagnosis and prognosis in cancer. *Int. J. Biol. Markers* 15, 84–89 (2000).
33. Golub, T.R. *et al.* Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science* 286, 531–537 (1999).
34. Hedenfalk, I. *et al.* Gene-expression profiles in hereditary breast cancer. *N. Engl. J. Med.* 344, 539–548 (2001).
35. Ohta, Y. *et al.* Vascular endothelial growth factor and lymph node metastasis in primary lung cancer. *Br. J. Cancer* 76, 1041–1045 (1997).
36. Shibusa, T., Shijubo, N. & Abe, S. Tumor angiogenesis and vascular endothelial growth factor expression in stage I lung adenocarcinoma. *Clin. Cancer Res.* 4, 1483–1487 (1998).
37. Girardin, S.E. & Yaniv, M. A direct interaction between JNK1 and Cdk1 is critical for Rac1-induced JNK activation. *EMBO J.* 20, 3437–3446 (2001).
38. Liu, E. *et al.* The Ras-mitogen-activated protein kinase pathway is critical for the activation of matrix metalloproteinase secretion and the invasiveness in v-*crk*-transformed 3Y1. *Cancer Res.* 60, 2361–64 (2000).
39. Hanahan, D. & Weinberg, R.A. The hallmarks of cancer. *Cell* 100, 57–70 (2000).
40. Hanson, L.A. *et al.* Expression of the glucocorticoid receptor and K-ras genes in urethan-induced mouse lung tumors and transformed cell lines. *Exp. Lung. Res.* 17, 371–387 (1991).
41. Lin, L. *et al.* A minimal critical region of the 8p22-23 amplicon in esophageal adenocarcinomas defined using STS-amplification mapping and quantitative PCR includes the GATA-4 gene. *Cancer Res.* 60, 1341–1347 (2000).
42. Kononen, J. *et al.* Tissue microarrays for high throughput molecular profiling of tumor specimens. *Nature Med.* 4, 844–847 (1998).
43. Johnson, R. & Wichern, D.W. *Applied Multivariate Statistical Analysis*. 543–578 (Prentice Hall, New Jersey, 1988).
44. Stone, M. Asymptotics for and against cross-validation. *Biometrika* 64, 29–38 (1977).
45. Cox, D.R. Regression models and life tables. *J.R. Stat. Soc.* 34, 187–220 (1972).





**Rapid quantitation of proinflammatory and chemoattractant cytokine expression in small tissue samples and monocyte-derived dendritic cells: validation of a new real-time RT-PCR technology.**

Blaschke V, Reich K, Blaschke S, Zipprich S, Neumann C.

Department of Dermatology, von-Siebold-Str. 3, D-37075, Goettingen, Germany.  
vblasch@gwdg.de

The analysis of cytokine profiles plays a central part in the characterization of disease-related inflammatory pathways and the identification of functional properties of immune cell subpopulations. Because tissue biopsy samples are too small to allow the detection of cytokine protein, the detection of mRNA by RT-PCR analysis is often used to investigate the cytokine milieu in inflammatory lesions. RT-PCR itself is a qualitative method, indicating the presence or absence of specific transcripts. With the use of internal or external standards it may also serve as a quantitative method. The most widely accepted method is quantitative competitive RT-PCR, based on internal shortened standards. Recently, online real-time PCR has been introduced (LightCycler), which allows quantitation in less than 30 min. Here, we have tested its use for the analysis of cytokine gene expression in different experimental *in vitro* and *ex vivo* settings. First, we compared quantitative competitive RT-PCR with real-time RT-PCR in the quantitation of transcription levels of the CD4(+) cell-specific chemoattractant Interleukin-16 during the maturation of monocyte-derived dendritic cells, and found a good correlation between both methods. Second, differences in the amounts of IL-16 mRNA in synovial tissue from patients with rheumatoid arthritis and osteoarthritis as assessed by real-time RT-PCR paralleled differences in the level of IL-16 protein in the synovial fluid. Finally, we employed real-time RT-PCR to study the cutaneous expression of several cytokines during experimental immunomodulatory therapy of psoriasis by Interleukin-10, and demonstrate that the technique is suitable for pharmacogenomic monitoring. In summary, real-time RT-PCR is a sensitive and rapid tool for quantifying mRNA expression even with small quantities of tissue. The results obtained do not differ from those generated by quantitative competitive RT-PCR.

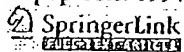
**Publication Types:**

- Evaluation Studies

PMID: 11121549 [PubMed - indexed for MEDLINE]

74: Apoptosis. 1997;2(6):518-28.

Related Articles, Links



**Butyrate-induced reversal of dexamethasone resistance in autonomous rat Nb2 lymphoma cells.**

Buckley AR, Krumenacker JS, Buckley DJ, Leff MA, Magnuson NS, Reed JC, Miyashita T, de Jong C, Gout PW.

Department of Pharmacology and Toxicology, University of North Dakota School of Medicine and Health Sciences, Grand Forks 58202-9037, USA.  
abuckley@mail.med.und.nodak.edu

The parental rat Nb2 lymphoma is a prolactin (PRL)-dependent T cell line. Exposure of a PRL-independent subline, Nb2-SFJCD1, to sodium butyrate (NaBT) causes transient reversal of their growth factor-independent proliferation in association with constitutive expression of protooncogenes pim-1 and c-myc. In the present study, we investigated the effect of NaBT treatment on the sensitivity of Nb2-SFJCD1 cells to dexamethasone (DEX)-induced apoptosis. Pretreatment with NaBT (2 mM, 72 h) partially reversed resistance to apoptosis in Nb2-SFJCD1 cells exposed to DEX (100 nM) for 12 h, assessed by flow cytometric analyses of DNA fragmentation. However, the cytolytic effect of DEX was abrogated by PRL in a time- and concentration-dependent manner. Evaluation of apoptosis-associated gene expression in NaBT-pre-treated cultures incubated with DEX or DEX+PRL indicated that the apoptosis resistance did not stem from altered bcl-2 or bax expression. However, there was a strong correlation between the resistance to DEX-activated apoptosis and their enhanced expression of pim-1 mRNA and protein. The results show that it is possible to reverse DEX-induced apoptosis of Nb2 pre-T cells and suggest the pim-1 gene product has an important role as a suppressor of this process, perhaps functioning as a mediator of PRL action.

PMID: 14646523 [PubMed]



283: Neurosci Lett. 1999 Apr 23;265(3):191-4.

[Related Articles](#), [Links](#)

ELSEVIER

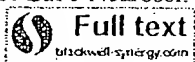
**Alterations in neuropeptide Y levels and Y1 binding sites in the Flinders Sensitive Line rats, a genetic animal model of depression.**

Caberlotto L, Jimenez P, Overstreet DH, Hurd YL, Mathe AA, Fuxe K.

Department of Neuroscience, Karolinska Institute, Stockholm, Sweden.

Previously, we observed specific alterations of neuropeptide Y (NPY) and Y1 receptor mRNA expression in discrete regions of the Flinders Sensitive Line rats (FSL), an animal model of depression. In order to clarify the correlation between mRNA expression and protein content, radioimmunoassay and receptor autoradiography were currently performed. In the FSL rats, NPY-like immunoreactivity (NPY-LI) was decreased in the hippocampal CA region, while Y1 binding sites were increased; NPY-LI was increased in the arcuate nucleus. Fluoxetine treatment elevated NPY-LI in the arcuate and anterior cingulate cortex and increased Y1 binding sites in the medial amygdala and occipital cortex in both strains. No differences were found regarding the Y2 binding sites. The results demonstrate a good correlation between NPY peptide and mRNA expression, and sustain the possible involvement of NPY and Y1 receptors in depression.

PMID: 10327163 [PubMed - indexed for MEDLINE]



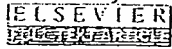
**Neurokinin 1 receptor and relative abundance of the short and long isoforms in the human brain.**

Caberlotto L, Hurd YL, Murdock P, Wahlin JP, Melotto S, Corsi M, Carletti R.

Department of Biology, Psychiatry CEDD, GlaxoSmithKline Medicine Research Centre, Verona, Italy. Laura.L.Caberlotto@gsk.com

Substance P exerts its various biochemical effects mainly via interactions through neurokinin-1 receptors (NK1). Recently, the NK1 receptor has attracted considerable interest for its possible role in a variety of psychiatric disorders including depression and anxiety. However, little is known regarding the anatomical distribution of NK1 in the human central nervous system (CNS). Riboprobe in situ hybridization, quantitative PCR and in vitro autoradiography were performed. Highest NK1 mRNA levels were localized in the locus coeruleus and ventral striatum, while moderate hybridization signals were observed in the cerebral cortex (most abundant in the visual cortex), hippocampus and different amygdaloid nuclei. Very low levels of the NK1 mRNA were detected in the cerebellum and thalamus. In view of the existence of a long and short isoform of the NK1 receptor, it was of interest to assess whether there was a differential distribution of the two splice variants in the human CNS and peripheral tissues. A quantitative TaqMan PCR analysis showed that the long NK1 isoform was the most prevalent throughout the human brain, while in peripheral tissues the truncated form was the most represented. <sup>3</sup>H-Substance P autoradiography revealed a good correlation between receptor binding sites and NK1 mRNA expression throughout the brain, with the highest levels of binding in the locus coeruleus. These results provide the anatomical evidence that the NK1 receptors have a strong association with neuronal systems relevant to mood regulation and stress in the human brain, but do not suggest a region-specific role of the two isoforms in the CNS.

PMID: 12752772 [PubMed - indexed for MEDLINE]



### Characterization of cyclin D2 expression in human endometrium.


Choi D, Yoon S, Lee E, Hwang S, Song S, Kim J, Yoon BK, Lee JH.

Department of Obstetrics and Gynecology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea. dschoi@smc.samsung.co.kr

**OBJECTIVE:** This study was undertaken to investigate cyclin D2 mRNA and protein expression in human endometrium during the menstrual cycle. **METHODS:** Endometrial samples were obtained from 15 premenopausal nonpregnant women who had hysterectomies for benign gynecologic reasons. They were divided into the following five groups according to histologic dating: early proliferative ( $n = 3$ ), mid to late proliferative ( $n = 3$ ), early secretory ( $n = 3$ ), mid secretory ( $n = 3$ ), and late secretory ( $n = 3$ ). Cyclin D2 mRNA and protein expression were analyzed using reverse transcriptase-polymerase chain reaction, Western blotting, and immunohistochemistry. **RESULTS:** Cyclin D2 mRNA and protein were expressed in human endometrial tissue throughout the menstrual cycle. Cyclin D2 mRNA and protein expression of proliferative phase endometrium were significantly higher than those of secretory phase endometrium ( $P < .05$ ). The staining intensity of cyclin D2 in proliferative phase endometrium was higher than that in secretory phase ( $P < .05$ ). Cyclin D2 mRNA level showed good correlation with cyclin D2 protein level ( $R = 0.579$ ,  $P < .03$ ), and cyclin D2 protein also showed good correlation with immunohistochemical staining intensity ( $R = 0.562$ ,  $P < .03$ ). **CONCLUSION:** Cyclin D2 was expressed in human endometrium throughout the menstrual cycle. Cyclin D2 mRNA and protein were expressed at high levels in proliferative phase endometrium, especially in the early proliferative phase, and then decreased in the secretory phase.

PMID: 11839508 [PubMed - indexed for MEDLINE]

45: Virchows Arch. 2004 Sep;445(3):279-84. Epub 2004 Aug 10. [Related Articles](#) [Links](#)

 [SpringerLink](#)  
[Database Links](#)

## Human chorionic gonadotrophin beta expression in malignant Barrett's oesophagus.

Couvelard A, Paraf F, Vidaud D, Dubois S, Vidaud M, Flejou JF, Degott C.

Service d'Anatomie Pathologique, Hopital Beaujon, 92118 Clichy cedex, France.  
[anne.couvelard@bjn.ap-hop-paris.fr](mailto:anne.couvelard@bjn.ap-hop-paris.fr)

**BACKGROUND:** Human chorionic gonadotrophin beta (hCGbeta) is expressed in several non-trophoblastic tumours, and this is usually associated with aggressive behaviour. Little is known about hCGbeta expression in Barrett's adenocarcinoma. **MATERIALS AND METHODS:** We determined the hCGbeta profile in a large series of surgically resected Barrett's adenocarcinoma (a) at mRNA level using real-time quantitative reverse-transcription polymerase chain reaction analysis and (b) at protein level using immunohistochemistry with a polyclonal antibody and with a monoclonal antibody specific for free hCGbeta. We then sought links between the hCGbeta protein expression pattern and clinical and pathological parameters, including patient outcome as well as vascular endothelial growth factor (VEGF) expression. **RESULTS:** hCGbeta protein expression was observed in 43 of 76 (57%) Barrett's adenocarcinomas. We showed a strong correlation between hCGbeta protein abundance and CGB mRNA level. We observed a statistical link between hCGbeta protein expression and infiltrative tumour type ( $P=0.023$ ), perineural neoplastic invasion ( $P=0.007$ ) and VEGF protein expression ( $P=0.016$ ). hCGbeta expression tended to be associated with a poor outcome (16% versus 36% survival 8 years after resection). **CONCLUSION:** Expression of hCGbeta correlates with specific infiltrative characteristics and is associated with higher VEGF expression. Both molecules may play a co-ordinated role in the development of Barrett's adenocarcinomas.

PMID: 15309632 [PubMed - indexed for MEDLINE]



**Downregulation of ENaC activity and expression by TNF-alpha in alveolar epithelial cells.**

Dagenais A, Fr  chette R, Yamagata Y, Yamagata T, Carmel JF, Clermont ME, Brochiero E, Masse C, Berthiaume Y.

Centre de recherche, CHUM-H  tel-Dieu, 3850 St-Urbain, Montreal, Quebec, Canada H2W 1T7. andre.dagenais.chum@ssss.gouv.qc.ca

Sodium absorption by an amiloride-sensitive channel is the main driving force of lung liquid clearance at birth and lung edema clearance in adulthood. In this study, we tested whether tumor necrosis factor-alpha (TNF-alpha), a proinflammatory cytokine involved in several lung pathologies, could modulate sodium absorption in cultured alveolar epithelial cells. We found that TNF-alpha decreased the expression of the alpha-, beta-, and gamma-subunits of epithelial sodium channel (ENaC) mRNA to 36, 43, and 16% of the controls after 24-h treatment and reduced to 50% the amount of alpha-ENaC protein in these cells. There was no impact, however, on alpha(1) and beta(1) Na(+)-K(+)-ATPase mRNA expression. Amiloride-sensitive current and ouabain-sensitive Rb(+) uptake were reduced, respectively, to 28 and 39% of the controls. A strong correlation was found at different TNF-alpha concentrations between the decrease of amiloride-sensitive current and alpha-ENaC mRNA expression. All these data show that TNF-alpha, a proinflammatory cytokine present during lung infection, has a profound influence on the capacity of alveolar epithelial cells to transport sodium.

PMID: 14514522 [PubMed - indexed for MEDLINE]

### **Involvement of the CCND1 gene in hairy cell leukemia.**

**de Boer CJ, Kluin-Nelemans JC, Dreef E, Kester MG, Kluin PM, Schuurin E, van Krieken JH.**

Department of Pathology, University of Leiden, The Netherlands.

**BACKGROUND:** Previous results suggested increased mRNA expression of CCND1 in hairy cell leukemia (HCL). The CCND1 gene is involved in the t(11;14)(q13;q32) chromosomal rearrangement, a characteristic abnormality in mantle cell lymphoma (MCL). We and others reported that, in contrast to other B-cell lymphomas, almost all MCL have over-expression of the CCND1 gene with a good correlation between RNA and protein analysis. Recent studies showed that overexpression of the cyclin D1 protein can be easily detected by immunohistochemistry (IHC) on formalin-fixed, paraffin embedded tissues. **PATIENTS AND METHODS:** To investigate whether the CCND1 gene is involved in HCL, we performed IHC on a series of 22 cases using formalin-fixed paraffin embedded splenectomy specimens. For IHC the sections were boiled in citrate buffer. The presence of rearrangements within the BCL-1 locus and the CCND1 gene was analyzed in 13 of 22 cases by Southern blot analysis using all available break-point probes. Expression of CCND1 was analyzed at the mRNA level (Northern blot) and protein level (IHC). **RESULTS:** Overexpression of the cyclin D1 protein using IHC was observed in all cases, with strong expression in 5 cases. Pre-existing B- and T-cell areas of the spleen did not express significant levels of the cyclin D1 protein. Seven of 9 cases analyzed by both IHC and Northern blotting showed overexpression of the CCND1 gene with both methods. No genomic abnormalities were observed in any of the 13 cases studied by Southern blot analysis. Additionally, no 11q13 abnormalities were detected by banding analysis of 19 of 22 cases. **CONCLUSIONS:** The elevated levels of CCND1 mRNA and protein in conjunction with the absence of overt rearrangements within the BCL-1 locus distinguish HCL from MCL and other B-cell malignancies. This suggests that activation of the CCND1 gene in HCL is due to mechanisms other than chromosomal rearrangement.

PMID: 8740788 [PubMed - indexed for MEDLINE]



**Inhibin and activin production and subunit expression in human placental cells cultured in vitro.**

**Debieve F, Pampfer S, Thomas K**

Department of Obstetrics and Gynecological Endocrinology, Universite Catholique de Louvain, 1200 Brussels, Belgium.

Inhibins and activins are dimeric proteins, with each subunit being one of three related protein subunits (alpha, betaA or betaB). The mRNA levels of these subunits were studied quantitatively during in-vitro differentiation of human cytotrophoblast cells into syncytium, using Northern blot analysis and semi-quantitative reverse transcription-polymerase chain reaction (RT-PCR) analysis. The corresponding protein concentrations were determined by specific enzyme-linked immunosorbent assays for inhibin A, B, pro alphaC and activin A in cellular protein extracts and culture medium (n = 5). Immunofluorescence studies showed syncytium formation after 48 h. The alpha subunit was present before plating and increased at 48 h ( $P < 0.001$ ) while the betaA subunit was weak before plating and increased at 24 h. The betaB subunit was not detected. With respect to corresponding protein synthesis, inhibin A (alpha + betaA) had risen after 48 h in cellular protein extract and after 72 h in culture medium, while activin A (betaA + betaB) was detected after 24 h, with no significant variations in culture medium. There was a good correlation between inhibin A and alpha subunit expression ( $r = 0.736$ ,  $P < 0.001$ ), as well as between activin A and betaA subunit expression ( $r = 0.755$ ,  $P < 0.001$ ). This study showed that mRNA expression parallels protein synthesis of inhibin and activin in trophoblast cells. Inhibin A synthesis appears to be dependent on alpha subunit mRNA expression, rather than on the betaA subunit which controls activin A synthesis. This study has also shown that isolated cytotrophoblast cells do not produce dimeric inhibin. However, during the transformation of cytotrophoblast cells into syncytium, betaA subunit mRNA expression may be an indicator of cell aggregation, while alpha subunit mRNA expression may be an indicator of cell fusion.

PMID: 10908285 [PubMed - indexed for MEDLINE]



**Expression of membrane-type matrix metalloproteinases 4, 5, and 6 in mouse corneas infected with *P. aeruginosa*.**

Dong Z, Katar M, Alousi S, Berk RS.

Department of Immunology and Microbiology, Wayne State University School of Medicine, 540 E. Canfield, Detroit, MI 48201, USA.

**PURPOSE:** To investigate the expression and regulation of membrane-type matrix metalloproteinases (MT-MMPs) 4, 5, and 6 in the mouse corneas infected with *Pseudomonas aeruginosa*. **METHODS:** C57BL/6J mice were intracorneally infected with *P. aeruginosa*. The expression of MT4-, MT5-, and MT6-MMP was detected at both the mRNA and protein levels by RT-PCR and immunoblot analysis. Immunohistochemical staining was performed to localize the expression of MT4- and MT5-MMP in the mouse corneas. **RESULTS:** Expression of MT4- and MT5-MMP was detected in the normal (uninfected) cornea by RT-PCR and immunoblot analysis. When infected with *P. aeruginosa*, the corneas showed significant induction of each MT-MMP. Localization of MT4- and MT5-MMP revealed that the expression of MT5-MMP was restricted to the epithelial tissue in the normal cornea, whereas the induced expression of MT4- and MT5-MMP was predominantly in the substantia propria, which contained most of the infiltrating cells. MT6-MMP expression was not detected in the uninfected cornea but was upregulated in the infected corneas. **CONCLUSIONS:** Expression of MT4-, MT5-, and MT6-MMP was induced in corneas infected with *P. aeruginosa*. Immunohistochemistry showed predominant immunoreactivity of MT4- and MT5-MMP in the substantia propria. Previous histologic studies have revealed different patterns of inflammatory cell infiltration with an increased number of polymorphonuclear neutrophils (PMNs) during the early stage of inflammation and increased macrophages during the late stage. These results indicate a good correlation between the overexpression of the MT-MMPs in the infected corneas and the inflammatory response—that is, leukocyte infiltration—indicating that inflammatory cells such as macrophages and PMNs may play a role in the upregulation of MT-MMPs during corneal infection, which in turn can cause the destruction of corneal tissue.

PMID: 11726626 [PubMed - indexed for MEDLINE]



**Assessment of proliferative activity in colorectal carcinomas by quantitative reverse transcriptase-polymerase chain reaction (RT-PCR).**

**Duchrow M, Hasemeyer S, Broll R, Bruch HP, Windhovel U.**

Surgical Research Laboratory, Surgical Clinic, Medical University of Lubeck, Ratzeburger Allee 160, D-23538 Lubeck, Germany.

The monoclonal antibody Ki-67 and the isospecific monoclonal antibody MIB-1 are routinely used in oncology to assess the proliferation index of tumor cells. A more objective and sensitive method is the determination of the of Ki-67 protein-specific mRNA by quantitative reverse transcriptase-polymerase chain reaction (RT-PCR). In 25 resected colorectal adenocarcinomas of different stages and grades we determined between 0.2 and 4.4 amol (10(-18) mol) Ki-67 protein-specific mRNA per microgram total RNA (median = 0.88 amol). The corresponding Ki-67 indices (expressing the percentage of Ki-67/MIB-1 positive tumor cells) ranged from 41 to 81% (median = 61%). We found a good correlation between Ki-67 index and mRNA expression ( $r = 0.75$ ), a significant correlation between both data and tumor stage (primary tumor, regional nodes, metastasis [pTNM] staging classification) ( $p < 0.001$ ), but not between both data and tumor grade. Both Ki-67 indices ( $p = 0.05$ ) and mRNA levels ( $p = 0.014$ ) correlated significantly to the patients' survival. These results demonstrate that the Ki-67 protein-specific quantitative RT-PCR is a useful method for the characterization of tumor cell proliferation.

PMID: 11486701 [PubMed - indexed for MEDLINE]

Comment in:

- [Equine Vet J. 2002 Jul;34\(4\):326-7.](#)

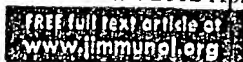
**Molecular characterisation of carbohydrate digestion and absorption in equine small intestine.**

Dyer J, Fernandez-Castano Merediz E, Salmon KS, Proudman CJ, Edwards GB, Shirazi-Beechey SP.

Department of Veterinary Preclinical Sciences, University of Liverpool, UK.

Dietary carbohydrates, when digested and absorbed in the small intestine of the horse, provide a substantial fraction of metabolisable energy. However, if levels in diets exceed the capacity of the equine small intestine to digest and absorb them, they reach the hindgut, cause alterations in microbial populations and the metabolite products and predispose the horse to gastrointestinal diseases. We set out to determine, at the molecular level, the mechanisms, properties and the site of expression of carbohydrate digestive and absorptive functions of the equine small intestinal brush-border membrane. We have demonstrated that the disaccharidases sucrase, lactase and maltase are expressed diversely along the length of the intestine and D-glucose is transported across the equine intestinal brush-border membrane by a high affinity, low capacity, Na<sup>+</sup>/glucose cotransporter type 1 isoform (SGLT1). The highest rate of transport is in duodenum > jejunum > ileum. We have cloned and sequenced the cDNA encoding equine SGLT1 and alignment with SGLT1 of other species indicates 85-89% homology at the nucleotide and 84-87% identity at the amino acid levels. We have shown that there is a good correlation between levels of functional SGLT1 protein and SGLT1 mRNA abundance along the length of the small intestine. This indicates that the major site of glucose absorption in horses maintained on conventional grass-based diets is in the proximal intestine, and the expression of equine intestinal SGLT1 along the proximal to distal axis of the intestine is regulated at the level of mRNA abundance. The data presented in this paper are the first to provide information on the capacity of the equine intestine to digest and absorb soluble carbohydrates and has implications for a better feed management, pharmaceutical intervention and for dietary supplementation in horses following intestinal resection.

PMID: 12117106 [PubMed - indexed for MEDLINE]



**Suppressors of cytokine signaling proteins are differentially expressed in Th1 and Th2 cells: implications for Th cell lineage commitment and maintenance.**

Egwuagu CE, Yu CR, Zhang M, Mahdi RM, Kim SJ, Gery I.

Laboratory of Immunology, National Eye Institute, National Institutes of Health, Bethesda, MD 20892, USA. emeka@helix.nih.gov

Positive regulatory factors induced by IL-12/STAT4 and IL-4/STAT6 signaling during T cell development contribute to polarized patterns of cytokine expression manifested by differentiated Th cells. These two critical and antagonistic signaling pathways are under negative feedback regulation by a multimember family of intracellular proteins called suppressor of cytokine signaling (SOCS). However, it is not known whether these negative regulatory factors also modulate Th1/Th2 lineage commitment and maintenance. We show here that CD4(+) naive T cells constitutively express low levels of SOCS1, SOCS2, and SOCS3 mRNAs. These mRNAs and their proteins increase significantly in nonpolarized Th cells after activation by TCR signaling. We further show that differentiation into Th1 or Th2 phenotype is accompanied by preferential expression of distinct SOCS mRNA transcripts and proteins. SOCS1 expression is 5-fold higher in Th1 than in Th2 cells, whereas Th2 cells contain 23-fold higher levels of SOCS3. We also demonstrate that IL-12-induced STAT4 activation is inhibited in Th2 cells that express high levels of SOCS3 whereas IL-4/STAT6 signaling is constitutively activated in Th2 cells, but not Th1 cells, with high SOCS1 expression. These results suggest that mutually exclusive use of STAT4 and STAT6 signaling pathways by differentiated Th cells may derive in part, from SOCS3- or SOCS1-mediated repression of IL-12/STAT4- or IL-4/STAT6 signaling in Th2 and Th1 cells, respectively. Given the strong correlation between distinct patterns of SOCS expression and differentiation into the Th1 or Th2 phenotype, SOCS1 and SOCS3 proteins are therefore Th lineage markers that can serve as therapeutic targets for immune modulation therapy.

PMID: 11907070 [PubMed - indexed for MEDLINE]

FREE full text article at  
bio.bmjjournals.com

## **Intravitreal invading cells contribute to vitreal cytokine milieu in proliferative vitreoretinopathy.**

**El-Ghrably IA, Dua HS, Orr GM, Fischer D, Tighe PJ.**

Larry A Donoso Laboratory for Eye Research, Department of Ophthalmology, University of Nottingham, UK.

**AIM:** To examine the contribution of infiltrating cells in the local production of cytokines within the vitreous of patients with proliferative vitreoretinopathy (PVR). **METHODS:** The presence of mRNA coding for IL-6, IL-8, IL-1beta, IL-1alpha, TNFalpha, IFNgamma, IL-12, and HPRT was investigated in 25 vitreous samples from patients with PVR, 11 vitreous samples from patients with retinal detachment (RD) not complicated by PVR, and 10 vitreous samples from patients with macular hole (MH). A quantitative reverse transcriptase polymerase chain reaction (RT-PCR) using an internal competitor was used to investigate these samples. From these samples, 15 PVR, 8 RD, and 8 MH were analysed for the protein levels of the same cytokines using enzyme linked immunosorbent assay (ELISA). Spearman correlation was used to test any association between mRNA and cytokine protein levels as an indicator of the contribution these cells make to the intravitreal cytokine milieu. **RESULTS:** A strong correlation was found between mRNA and their respective cytokine levels (protein products) for IL-6, IL-8, IL-1beta, IL-1alpha, TNFalpha, IFNgamma (Spearman  $r = 0.83, 0.73, 0.67, 0.91, 0.73,$  and  $0.73$  respectively), but not for IL-12. The median levels of IL-6, IL-8, IL-1beta, and IFNgamma mRNA and their respective cytokines were significantly higher ( $p < 0.05$ ) in patients with PVR than in those with macular hole. There was no statistically significant difference in the median levels of IL-1alpha mRNA between PVR and MH but the cytokine IL-1alpha was detected at a significantly higher level in PVR compared with MH patients. Between PVR and RD patients, there was no statistically significant difference in mRNA levels for all the investigated cytokines ( $p > 0.05$ ) except for IL-6 where there was a statistical significance ( $p = 0.038$ ). In contrast, the median levels of IL-6, IL-8, and IL-1beta cytokines were significantly higher ( $p < 0.05$ ) in patients with PVR than in those with RD, whereas for IL-1alpha and IFNgamma no significant statistical difference was detected between PVR and RD patients ( $p > 0.05$ ). When results of RD and MH patients were compared, a statistical difference was only detected in mRNA levels of IFNgamma ( $p = 0.008$ ). However, no difference was detected for IFNgamma (protein product) or for any of the other cytokines between RD and MH patients. **CONCLUSION:** Levels of both protein and mRNA encoding IL-6, IL-8, IL-1beta, and IFNgamma is significantly increased in vitreous samples from patients with PVR. The strong correlation between ELISA detectable cytokines (protein products) and their respective mRNA levels suggest that intravitreal, invasive cells are the major source of these cytokines, with the exception of IL-12. Cells invading the vitreous do not appear to locally produce IL-12 mRNA. This would appear to implicate cells peripheral to the

vitreal mass as the major source of this cytokine.

PMID: 11264138 [PubMed - indexed for MEDLINE]

## Modulation of the glutamatergic receptors (AMPA and NMDA) and of glutamate vesicular transporter 2 in the rat facial nucleus after axotomy.

Eleore L, Vassias I, Vidal PP, de Waele C.

LNRS (CNRS-Paris V), ESA 7060, Centre Universitaire des Saints-Peres, 45 rue des Saints-Peres, 75270 Paris Cedex 06, France.

Facial nerve axotomy is a good model for studying neuronal plasticity and regeneration in the peripheral nervous system. We investigated in the rat the effect of axotomy on the different subunits of excitatory glutamatergic AMPA (GluR1-4), NMDA (NR1, NR2A-D) receptors, post-synaptic density 95, vesicular glutamate transporter 2, beta catenin and cadherin. mRNA levels and/or protein production were analyzed 1, 3, 8, 30 and 60 days after facial nerve axotomy by in situ hybridization and immunohistofluorescence. mRNAs coding for the GluR2-4, NR1, NR2A, B, D subunits of glutamatergic receptors and for post-synaptic density 95, were less abundant after axotomy. The decrease began as early as 1 or 3 days after axotomy; the mRNAs levels were lowest 8 days post-lesion, and returned to normal or near normal 60 days after the lesion. The NR2C subunit mRNAs were not detected in either lesioned or intact facial nuclei. Immunohistochemistry using specific antibodies against GluR2-3 subunits and against NR1 confirmed this down-regulation. There was also a large decrease in vesicular glutamate transporter 2 immunostaining in the axotomized facial nuclei at early stages following facial nerve section. In contrast, no decrease of NR2A subunit and of post-synaptic density 95 could be detected at any time following the lesion. beta Catenin and cadherin immunoreactivity pattern changed around the cell body of facial motoneuron by day 3 after axotomy, and then, tends to recover at day post-lesion 60 days. Therefore, our results suggest a high correlation between restoration of nerve/muscle synaptic contact, synaptic structure and function in facial nuclei. To investigate the mechanisms involved in the change of expression of these proteins following axotomy, the facial nerve was perfused with tetrodotoxin for 8 days. The blockade of action potential significantly decreased GluR2-3, NR1 and NR2A mRNAs in the ipsilateral facial nuclei. Thus, axotomy-induced changes in mRNA abundance seemed to depend partly on disruption of activity.

PMID: 16182453 [PubMed - in process]

**Altered levels of scavenging enzymes in embryos subjected to a diabetic environment.**

Forsberg H, Borg LA, Cagliero E, Eriksson UJ.

Department of Medical Cell Biology, University of Uppsala, Sweden.

Maternal diabetes during pregnancy is associated with an increased rate of congenital malformations in the offspring. The exact molecular etiology of the disturbed embryogenesis is unknown, but an involvement of radical oxygen species in the teratological process has been suggested. Oxidative damage presupposes an imbalance between the activity of the free oxygen radicals and the antioxidant defence mechanisms on the cellular level. The aim of the present study was to investigate if maternal diabetes *in vivo*, or high glucose *in vitro* alters the expression of the free oxygen radical scavenging enzymes superoxide dismutase (CuZnSOD and MnSOD), catalase and glutathione peroxidase in rat embryos during late organogenesis. We studied offspring of normal and diabetic rats on gestational days 11 and 12, and also evaluated day-11 embryos after a 48 hour culture period in 10 mM or 50 mM glucose concentration. Both maternal diabetes and high glucose culture caused growth retardation and increased rate of congenital malformations in the embryos. The CuZnSOD and MnSOD enzymes were expressed on gestational day 11 and both CuZnSOD, MnSOD and catalase were expressed on day 12 with increased concentrations of MnSOD transcripts when challenged by a diabetic milieu. There was a good correlation between mRNA, protein, and activity levels, suggesting that the regulation of these enzymes occurs primarily at the pretranslational level. Maternal diabetes *in vivo* and high glucose concentration *in vitro* induced increased MnSOD expression, concomitant with increased total SOD activity, and a tentative decrease in catalase expression and activity in the embryos. These findings support the notion of enhanced oxidative stress in the embryo as an etiologic agent in diabetic teratogenesis.

PMID: 8804988 [PubMed - indexed for MEDLINE]

Full text article at  
[endo.oxfordjournals.org](http://endo.oxfordjournals.org)

**Induction of the estrogen receptor by growth hormone and glucocorticoid substitution in primary cultures of rat hepatocytes.**

Freyschuss B, Stavreus-Evers A, Sahlin L, Eriksson H.

Department of Reproductive Endocrinology, Karolinska Hospital, Stockholm, Sweden.

Hepatic estrogen receptors (ER) mediate estrogenic effects on mammalian liver metabolism and are thereby involved in the regulation of important physiological/pathological processes, such as coagulation, atherosclerosis, and hypertension. The regulation of the formation of the ER in primary cultures of rat hepatocytes was studied by assaying ER and ER mRNA under different endocrine conditions. The ER concentration was measured using two different methods, a ligand-binding technique and an ER enzyme immunoassay. The results obtained by the two methods showed good correlation, and linear regression analysis gave a correlation coefficient of 0.95. ER concentrations fell to low steady state levels within 16 h after establishing the cell culture and remained low in the absence of hormonal substitution. Upon medium supplementation with pituitary GH and the glucocorticoid dexamethasone (DEX) in combination, the ER concentration increased 6-fold from  $4.2 \pm 1.0$  to  $25.8 \pm 7.0$  fmol/mg cytosolic protein. ER mRNA was measured by solution hybridization. Substitution with GH and DEX in combination increased ER mRNA to  $210 \pm 14\%$  of control levels. No effect on ER mRNA stability was seen after hormone treatment. It is concluded that the regulatory effects of GH and DEX on the hepatic ER in this in vitro system are very similar to the effects of these hormones under in vivo conditions. The inducible expression of the ER has never before, to our knowledge, been demonstrated in any mammalian liver cell culture system.

PMID: 8404593 [PubMed - indexed for MEDLINE]



1: *Blood*. 2005 Dec 15;106(13):4315-21. Epub 2005 Aug 25.

Related Articles, Links

Full text article at  
[www.bloodjournal.org](http://www.bloodjournal.org)

**Cyclin D1-negative mantle cell lymphoma: a clinicopathologic study based on gene expression profiling.**

Fu K, Weisenburger DD, Greiner TC, Dave S, Wright G, Rosenwald A, Chiorazzi M, Iqbal J, Gesk S, Siebert R, De Jong D, Jaffe ES, Wilson WH, Delabie J, Ott G, Dave BJ, Sanger WG, Smith LM, Rimsza L, Braziel RM, Muller-Hermelink HK, Campo E, Gascoyne RD, Staudt LM, Chan WC; Lymphoma/Leukemia Molecular Profiling Project.

Department of Pathology and Microbiology, University of Nebraska Medical Center, 983135 Nebraska Medical Center, Omaha, NE 68198-3135, USA. [kfu@unmc.edu](mailto:kfu@unmc.edu)

Cyclin D1 overexpression is believed to be essential in the pathogenesis of mantle cell lymphoma (MCL). Hence, the existence of cyclin D1-negative MCL has been controversial and difficult to substantiate. Our previous gene expression profiling study identified several cases that lacked cyclin D1 expression, but had a gene expression signature typical of MCL. Herein, we report the clinical, pathologic, and genetic features of 6 cases of cyclin D1-negative MCL. All 6 cases exhibited the characteristic morphologic features and the unique gene expression signature of MCL but lacked the t(11;14)(q13; q32) by fluorescence in situ hybridization (FISH) analysis. The tumor cells also failed to express cyclin D1 protein, but instead expressed either cyclin D2 (2 cases) or cyclin D3 (4 cases). There was good correlation between cyclin D protein expression and the corresponding mRNA expression levels by gene expression analysis. Using interphase FISH, we did not detect chromosomal translocations or amplifications involving CCND2 and CCND3 loci in these cases. Patients with cyclin D1-negative MCL were similar clinically to those with cyclin D1-positive MCL. In conclusion, cases of cyclin D1-negative MCL do exist and are part of the spectrum of MCL. Up-regulation of cyclin D2 or D3 may substitute for cyclin D1 in the pathogenesis of MCL.

PMID: 16123218 [PubMed - indexed for MEDLINE]



**Oxytocin receptors in bovine cervix: distribution and gene expression during the estrous cycle.**

Fuchs AR, Ivell R, Fields PA, Chang SM, Fields MJ.

Department of Obstetrics and Gynecology, Cornell University Medical College, New York, New York 10021, USA.

Oxytocin (OT) receptor (OTR) concentrations were determined in the cervix of nonpregnant cows on cycle Days 0, 3, 7-8, 17, and 19 (n = 3-4 cows each day); [<sup>3</sup>H]OT was used as the labeled ligand. Mucosal and muscle layers of the cervix were also analyzed separately for both ligand binding and expression of the OTR gene using a newly developed RNase protection assay (RAP). Cellular localization of OTR protein was determined by immunohistochemistry. All regions of cervix from cows at estrus had high concentrations of OTR; in the luteal phase, all were sharply down-regulated. At estrus the mucosal layer had about 30-fold higher concentrations than the muscle layer. OTR mRNA was readily detected by RAP in the mucosa from estrous cows, while much weaker signals were found in the muscle. On Days 7-17, the OTR mRNA signals in both mucosa and muscle were very faint or nondetectable. Thus, there was a good correlation between ligand binding and mRNA expression, which suggests that OTR concentrations are mainly regulated at the transcriptional level. The epithelial cells at the luminal surface of the mucosa were the principal site of immunoreactive OTR; muscle cells showed significantly weaker signals. Previously, OT was found to stimulate prostaglandin (PG) E<sub>2</sub> output in vitro in bovine cervical tissues. Since PGE<sub>2</sub> is capable of softening the cervix, our findings suggest that OT may have a novel physiological function to cause softening of the bovine cervix mediated by the release of PGE<sub>2</sub>.

PMID: 8835394 [PubMed - indexed for MEDLINE]

### **Silencing of the thrombomodulin gene in human malignant melanoma.**

**Furuta J, Kaneda A, Umehayashi Y, Otsuka F, Sugimura T, Ushijima T.**

Carcinogenesis Division, National Cancer Center Research Institute, Tokyo, Japan.

The loss of thrombomodulin (TM) expression is associated with tumour growth, infiltration and lymph node metastasis in human tumours. In melanoma cell lines, TM is reported to mediate cell adhesion, and its introduction into TM-negative melanoma cell lines suppresses their growth. In this study, we analysed TM expression in surgical melanoma specimens and the role of its promoter methylation in the loss of its expression. In 15 (75%) of the 20 specimens (five from a primary site and 15 from metastatic sites), melanoma cells lacked TM immunoreactivity. Methylation of the TM promoter region was detected in 10 (67%) of the 15 TM-negative specimens by methylation-specific polymerase chain reaction, whereas methylation was detected in two (40%) of the five TM-positive specimens. In cell lines, complete methylation of the TM promoter CpG island was detected in six (46%) of 13 melanoma cell lines, whereas no methylation was detected in two cultured normal melanocytes. There was a good correlation between the methylated status of the CpG island and the loss of TM messenger RNA (mRNA) expression. Treatment of melanoma cell lines with a demethylating agent, 5-aza-2'-deoxycytidine, induced demethylation of the promoter CpG island and the restoration of mRNA and protein expression. These findings suggest that most human melanomas lack TM expression, and that methylation of the promoter CpG island is one of the mechanisms responsible.

PMID: 15714116 [PubMed - indexed for MEDLINE]

266: Mol Cell Biol. 1999 Nov;19(11):7357-68.

[Related Articles](#), [Links](#)

FREE full text article on  
mcb.asm.org

Full text article  
in PubMed Central

## A sampling of the yeast proteome.

Futcher B, Latter GI, Monardo P, McLaughlin CS, Garrels JI.

Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA.  
[futcher@cshl.org](mailto:futcher@cshl.org)

In this study, we examined yeast proteins by two-dimensional (2D) gel electrophoresis and gathered quantitative information from about 1,400 spots. We found that there is an enormous range of protein abundance and, for identified spots, a good correlation between protein abundance, mRNA abundance, and codon bias. For each molecule of well-translated mRNA, there were about 4,000 molecules of protein. The relative abundance of proteins was measured in glucose and ethanol media. Protein turnover was examined and found to be insignificant for abundant proteins. Some phosphoproteins were identified. The behavior of proteins in differential centrifugation experiments was examined. Such experiments with 2D gels can give a global view of the yeast proteome.

PMID: 10523624 [PubMed - indexed for MEDLINE]

## A Sampling of the Yeast Proteome

B. FUTCHER,<sup>1\*</sup> G. I. LATTER,<sup>1</sup> P. MONARDO,<sup>1</sup> C. S. McLAUGHLIN,<sup>2</sup> AND J. I. GARRELS<sup>3</sup>

<sup>1</sup>Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724<sup>1</sup>; <sup>2</sup>Department of Biological Chemistry, University of California, Irvine, California 92717<sup>2</sup>; and <sup>3</sup>Proteome, Inc., Beverly, Massachusetts 01915<sup>3</sup>

Received 15 June 1999/Returned for modification 16 July 1999/Accepted 28 July 1999

In this study, we examined yeast proteins by two-dimensional (2D) gel electrophoresis and gathered quantitative information from about 1,400 spots. We found that there is an enormous range of protein abundance and, for identified spots, a good correlation between protein abundance, mRNA abundance, and codon bias. For each molecule of well-translated mRNA, there were about 4,000 molecules of protein. The relative abundance of proteins was measured in glucose and ethanol media. Protein turnover was examined and found to be insignificant for abundant proteins. Some phosphoproteins were identified. The behavior of proteins in differential centrifugation experiments was examined. Such experiments with 2D gels can give a global view of the yeast proteome.

The sequence of the yeast genome has been determined (9). More recently, the number of mRNA molecules for each expressed gene has been measured (27, 30). The next logical level of analysis is that of the expressed set of proteins. We have begun to analyze the yeast proteome by using two-dimensional (2D) gels.

2D gel electrophoresis separates proteins according to isoelectric point in one dimension and molecular weight in the other dimension (21), allowing resolution of thousands of proteins on a single gel. Although modern imaging and computing techniques can extract quantitative data for each of the spots in a 2D gel, there are only a few cases in which quantitative data have been gathered from 2D gels. 2D gel electrophoresis is almost unique in its ability to examine biological responses over thousands of proteins simultaneously and should therefore allow us a relatively comprehensive view of cellular metabolism.

We and others have worked toward assembling a yeast protein database consisting of a collection of identified spots in 2D gels and of data on each of these spots under various conditions (2, 7, 8, 10, 23, 25). These data could then be used in analyzing a protein or a metabolic process. *Saccharomyces cerevisiae* is a good organism for this approach since it has a well-understood physiology as well as a large number of mutants, and its genome has been sequenced. Given the sequence and the relative lack of introns in *S. cerevisiae*, it is easy to predict the sequence of the primary protein product of most genes. This aids tremendously in identifying these proteins on 2D gels.

There are three pillars on which such a database rests: (i) visualization of many protein spots simultaneously, (ii) quantification of the protein in each spot, and (iii) identification of the gene product for each spot. Our first efforts at visualization and identification for *S. cerevisiae* have been described elsewhere (7, 8). Here we describe quantitative data for these proteins under a variety of experimental conditions.

### MATERIALS AND METHODS

**Strains and media.** *S. cerevisiae* W303 (MATa *ade2-1 his3-11,15 leu2-3,112 trp1-1 ura3-1 can1-100*) was used (26). -Met YNB (yeast nitrogen base) medium was 1.7 g of YNB (Difco) per liter, 5 g of ammonium sulfate per liter, and

adenine, uracil, and all amino acids except methionine; -Met, -Cys YNB medium was the same but without methionine or cysteine. Medium was supplemented with 2% glucose (for most experiments) or with 2% ethanol (for ethanol experiments). Low-phosphate YEPD was described by Warner (28).

**Isotopic labeling of yeast and preparation of cell extracts.** Yeast strains were labeled and proteins were extracted as described by Garrels et al. (7, 8). Briefly, cells were grown to  $5 \times 10^6$  cells per ml at 30°C; 1 ml of culture was transferred to a fresh tube, and 0.3 mCi of [<sup>35</sup>S]methionine (e.g., Express protein labeling mix; New England Nuclear) was added to this 1-ml culture. The cells were incubated for a further 10 to 15 min and then transferred to a 1.5-ml microcentrifuge tube, chilled on ice, and harvested by centrifugation. The supernatant was removed, and the cell pellet was resuspended in 100  $\mu$ l of lysis buffer (20 mM Tris-HCl [pH 7.6], 10 mM NaF, 10 mM sodium pyrophosphate, 0.5 mM EDTA, 0.1% deoxycholate; just before use, phenylmethylsulfonyl fluoride was added to 1 mM, leupeptin was added to 1  $\mu$ g/ml, pepstatin was added to 1  $\mu$ g/ml, tosyl-sulfonyl phenylalanyl chloromethyl ketone was added to 10  $\mu$ g/ml, and soybean trypsin inhibitor was added to 10  $\mu$ g/ml).

The resuspended cells were transferred to a screw-cap 1.5-ml polypropylene tube containing 0.28 g of glass beads (0.5-mm diameter; Biospec Products) or 0.40 g of zirconia beads (0.5-mm diameter; Biospec Products). After the cap was secured, the tube was inserted into a MiniBeadbeater 8 (Biospec Products) and shaken at medium high speed at 4°C for 1 min. Breakage was typically 75%. Tubes were then spun in a microcentrifuge for 10 s at  $5,000 \times g$  at 4°C.

With a very fine pipette tip, liquid was withdrawn from the beads and transferred to a prechilled 1.5-ml tube containing 7  $\mu$ l of DNase I (0.5 mg/ml; Cooper product no. 6330)-RNase A (0.25 mg/ml; Cooper product no. 5679)-Mg (50 mM MgCl<sub>2</sub>) mix. Typically 70  $\mu$ l of liquid was recovered. The mixture was incubated on ice for 10 min to allow the RNase and DNase to work.

Next, 75  $\mu$ l of 2 $\times$  SDS (2 $\times$  SDS is 0.6% sodium dodecyl sulfate [SDS], 2% mercaptoethanol, and 0.1 M Tris-HCl [pH 8]) was added. The tube was plunged into boiling water, incubated for 1 min, and then plunged into ice. After cooling, the tube was centrifuged at 4°C for 3 min at  $14,000 \times g$ . The supernatant was transferred to a fresh tube and frozen at -70°C. About 5  $\mu$ l of this supernatant was used for each 2D gel.

**2D polyacrylamide gels.** 2D gels were made and run as described elsewhere (6-8).

**Image analysis of the gels.** The Quest II software system was used for quantitative image analysis (20, 22). Two techniques were used to collect quantitative data for analysis by Quest II software. First, before the advent of phosphorimagers, gels were dried and fluorographed. Each gel was exposed to film for three different times (typically 1 day, 2 weeks, and 6 weeks) to increase the dynamic range of the data. The films were scanned along with calibration strips to relate film optical density to disintegrations per minute in the gels and analyzed by the software to obtain a linear relationship between disintegrations per minute in the spots and optical densities of the film images. The quantitative data are expressed as parts per million of the total cellular protein. This value is calculated from the disintegrations per minute of the sample loaded onto the gel and by comparing the film density of each data spot with density of the film over the calibration strips of known radioactivity exposed to the same film. This yields the disintegrations per minute per millimeter for each spot on the gel and thence its parts-per-million value.

After the advent of phosphorimaging, gels bearing <sup>35</sup>S-labeled proteins were exposed to phosphorimager screens and scanned by a Fuji phosphorimager, typically for two exposures per gel. Calibration strips of known radioactivity were exposed simultaneously. Scan data from the phosphorimager was assimilated by Quest II software, and quantitative data were recorded for the spots on the gels.

\* Corresponding author. Mailing address: Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724. Phone: (516) 367-8828. Fax: (516) 367-8369. E-mail: fletcher@cshl.org.

**Measurements of protein turnover.** Cells in exponential phase were pulse-labeled with [ $^{35}$ S]methionine, excess cold Met and Cys were added, and samples of equal volume were taken from the culture at intervals up to 90 min (in one experiment) or up to 160 min (in a second experiment). Incorporation of  $^{35}$ S into protein was essentially 100% by the first sample (10 min). Extracts were made, and equal fractions of the samples were loaded on 2D gels (i.e., the different samples had different amounts of protein but equal amounts of  $^{35}$ S). Spots were quantitated with a phosphorimaging and Quest software.

The software was queried for spots whose radioactivity decreased through the time course. The algorithm examined all data points for all spots, drew a best-fit line through the data points, and looked for spots where this line had a statistically significant negative slope. In one of the experiments, there was one such spot. To the eye, this was a minor, unidentified spot seen only in the first two samples (10 and 20 min). In the other experiment, the Quest software found no spots meeting the criteria. Therefore, we concluded that none of the identified spots (and all but one of the visible spots) represented proteins with long half-lives.

**Centrifugal fractionation.** Cells were labeled, harvested, and broken with glass beads by the standard method described above except that no detergent (i.e., no deoxycholate) was present in the lysis buffer. The crude lysate was cleared of unbroken cells and large debris by centrifugation at  $300 \times g$  for 30 s. The supernatant of this centrifugation was then spun at  $16,000 \times g$  for 10 min to give the pellet used for Fig. 6B. The supernatant of the  $16,000 \times g$ , 10-min spin was then spun at  $100,000 \times g$  for 30 min to give the supernatant used for Fig. 6A.

**Protein abundance calculations.** A haploid yeast cell contains about  $4 \times 10^{12}$  g of protein (1, 15). Assuming a mean protein mass of 50 kDa, there are about  $50 \times 10^6$  molecules of protein per cell. There are about 1.8 methionines per 10 kDa of protein mass, which implies  $4.5 \times 10^8$  molecules of methionine per cell (neglecting the small pool of free Met). We measured (i) the counts per minute in each spot on the 2D gels, (ii) the total number of counts on each gel (by integrating counts over the entire gel), and (iii) the total number of counts loaded on the gel (by scintillation counting of the original sample). Thus, we know what fraction of the total incorporated radioactivity is present in each spot. After correcting for the methionine (and cysteine [see below]) content of each protein, we calculated an absolute number of protein molecules based on the fraction of radioactivity in each spot and on  $50 \times 10^6$  total molecules per cell.

The labeling mixture used contained about one-fifth as much radioactive cysteine as radioactive methionine. Therefore, the number of cysteine molecules per protein was also taken into account in calculating the number of molecules of protein, but Cys molecules were weighted one-fifth as heavily as Met molecules.

**mRNA abundance calculations.** For estimation of mRNA abundance, we used SAGE (serial analysis of gene expression) data (27) and Affymetrix chip hybridization data (29a, 30). The mRNA column in Table 1 shows mRNA abundance calculated from SAGE data alone. However, the SAGE data came from cells growing in YEPD medium, whereas our protein measurements were from cells growing in YNB medium. In addition, SAGE data for low-abundance mRNAs suffers from statistical variation. Therefore, we also used chip hybridization data (29a, 30) for mRNA from cells grown in YNB. These hybridization data also had disadvantages. First, the amounts of high-abundance mRNAs were systematically underestimated, probably because of saturation in the hybridizations, which used 10  $\mu$ g of cRNA. For example, the abundance of *ADH1* mRNA was 197 copies per cell by SAGE but only 32 copies per cell by hybridization, and the abundance of *ENO2* mRNA was 248 copies per cell by SAGE but only 41 by hybridization. When the amount of cRNA used in the hybridization was reduced to 1  $\mu$ g, the apparent amounts of mRNA were similar to the amounts determined by SAGE (29a, 29b). However, experiments using 1  $\mu$ g of cRNA have been done for only some genes (29a). Because amounts of mRNA were normalized to 15,000 per cell, and because the amounts of abundant mRNAs were underestimated, there is a 2.2-fold overestimate of the abundance of nonabundant mRNAs. We calculated this factor of 2.2 by adding together the number of mRNA molecules from a large number of genes expressed at a low level for both SAGE data and hybridization data. The sum for the same genes from hybridization data is 2.2-fold greater than that from SAGE data.

To take into account these difficulties, we compiled a list of "adjusted" mRNA abundance as follows. For all high-abundance mRNAs of our identified proteins, we used SAGE data. For all of these particular mRNAs, chip hybridization suggested that mRNA abundance was the same in YEPD and YNB media. For medium-abundance mRNAs, SAGE data were used, but when hybridization data showed a significant difference between YEPD and YNB, then the SAGE data were adjusted by the appropriate factor. Finally, for low-abundance mRNAs, we used data from chip hybridizations from YNB medium but divided by 2.2 to normalize to the SAGE results. These calculations were completed without reference to protein abundance.

**CAI.** The codon adaptation index (CAI) was taken from the yeast proteome database (YPD) (13), for which calculations were made according to Sharp and Li (24). Briefly, the index uses a reference set of highly expressed genes to assign a value to each codon, and then a score for a gene is calculated from the frequency of use of the various codons in that gene (24).

**Statistical analysis.** The JMP program was used with the aid of T. Tully. The JMP program showed that neither mRNA nor protein abundances were normally distributed; therefore, Spearman rank correlation coefficients ( $r_s$ ) were

calculated. The mRNA (adjusted and unadjusted) and protein data were also transformed so that Pearson product-moment correlation coefficients ( $r_p$ ) could be calculated. First, this was done by a Box-Cox transformation of log-transformed data. This transformation produced normal distributions, and an  $r_p$  of 0.76 was achieved. However, because the Box-Cox transformation is complex, we also did a simpler logarithmic transformation. This produced a normal distribution for the protein data. However, the distribution for the mRNA and adjusted mRNA data was close to, but not quite, normal. Nevertheless, we calculated the  $r_p$  and found that it was 0.76, identical to the coefficient from the Box-Cox transformed data. We therefore believe that this correlation coefficient is not misleading, despite the fact that the log(mRNA) distribution is not quite normal.

## RESULTS

**Visualization of 1,400 spots on three gel systems.** Yeast proteins have isoelectric points ranging from 3.1 to 12.8, and masses ranging from less than 10 kDa to 470 kDa. It is difficult to examine all proteins on a single kind of gel, because a gel with the needed range in pI and mass would give poor resolution of the thousands of spots in the central region of the gel. Therefore, we have used three gel systems: (i) pH "4 to 8" with 10% polyacrylamide; (ii) pH "3 to 10" with 10% polyacrylamide; and (iii) nonequilibrium with 15% polyacrylamide (7, 8). Each gel system allows good resolution of a subset of yeast proteins.

Figure 1 shows a pH 4–8, 10% polyacrylamide gel. The pH at the basic end of the isoelectric focusing gel cannot be maintained throughout focusing, and so the proteins resolved on such gels have isoelectric points between pH 4 and pH 6.7. For these pH 4–8 gels, we see 600 to 900 spots on the best gels after multiple exposures.

The pH 3–10 gels (not shown) extend the pI range somewhat beyond pH 7.5, allowing detection of several hundred additional spots. Finally, we use nonequilibrium gels with 15% acrylamide in the second dimension. These allow visualization of about 100 very basic proteins and about 170 small proteins (less than 20 kDa). In total, using all three gel systems, about 1,400 spots can be seen. These represent about 1,200 different proteins, which is about one-quarter to one-third of the proteins expressed under these conditions (27, 30). Here, we focus on the proteins seen on the pH 4–8 gels.

Although nearly all expressed proteins are present on these gels, the number seen is limited by a problem we call coverage. Since there are thousands of proteins on each gel, many proteins comigrate or nearly comigrate. When two proteins are resolved, but are close together, and one protein spot is much more intense than the other, a problem arises in visualizing the weaker spot: at long exposures when the weak signal is strong enough for detection, the signal from the strong spot spreads and covers the signal from the weaker spot. Thus, weak spots can be seen only when they are well separated from strong spots.

For a given gel, the number of detectable spots initially rises with exposure time. However, beyond an optimal exposure, the number of distinguishable spots begins to decrease, because signals from strong spots cover signals from nearby weak spots. At long exposures, the whole autoradiogram turns black. Thus, there is an optimum exposure yielding the maximum number of spots, and at this exposure the weakest spots are not seen.

Largely because of the problem of coverage, the proteins seen are strongly biased toward abundant proteins. All identified proteins have a CAI of 0.18 or more, and we have identified no transcription factors or protein kinases, which are nonabundant proteins. Thus, this technology is useful for examining protein synthesis, amino acid metabolism, and glycolysis but not for examining transcription, DNA replication, or the cell cycle.



**Spot identification.** The identification of various spots has been described elsewhere (7, 8). At present, 169 different spots representing 148 proteins have been identified. Many of these spots have been independently identified (2, 10, 23, 25). The main methods used in spot identification have been analysis of amino acid composition, gene overexpression, peptide sequencing, and mass spectrometry.

**Pulse-chase experiments and protein turnover.** Pulse-chase experiments were done to measure protein half-lives (Materials and Methods). Cells were labeled with [ $^{35}$ S]methionine for 10 min, and then an excess of unlabeled methionine was added. Samples were taken at 0, 10, 20, 30, 60, and 90 min after the beginning of the chase. Equal amounts of  $^{35}$ S were loaded from each sample; 2D gels were run, and spots were quantitated. Surprisingly, almost every spot was nearly constant in amount of radioactivity over the entire time course (not shown). A few spots shifted from one position to another because of post-translational modifications (e.g., phosphorylation of Rpa0 and Efb1). Thus, the proteins being visualized are all or nearly all very stable proteins, with half-lives of more than 90 min. Gygi et al. (10) have come to a similar conclusion by using the N-end rule to predict protein half-lives. This result does not imply that all yeast proteins are stable. The proteins being visualized are abundant proteins; this is partly because they are stable proteins.

**Protein quantitation.** Because all of the proteins seen had effectively the same half-life, the abundance of each protein was directly proportional to the amount of radioactivity incorporated during labeling. Thus, after taking into account the total number of protein molecules per cell, the average content of methionine and cysteine, and the methionine and cysteine content of each identified protein, we could calculate the abundance of each identified protein (Tables 1 and 2; Materials and Methods). About 1,000 unidentified proteins were also quantified, assuming an average content of Met and Cys.

Many proteins give multiple spots (7, 8). The contribution from each spot was summed to give the total protein amount. However, many proteins probably have minor spots that we are not aware of, causing the amount of protein to be underestimated.

When the proteins on a pH 4–8 gel were ordered by abundance, the most abundant protein had 8,904 ppm, the 10th most abundant had 2,842 ppm, the 100th most abundant had 314 ppm, the 500th most abundant had 57 ppm, and the 1,000th most abundant (visualized at greater than optimum exposure) had 23 ppm. Thus, there is more than a 300-fold range in abundance among the visualized proteins. The most abundant 10 proteins account for about 25% of the total protein on the pH 4–8 gel, the most abundant 60 proteins account for 50%, and the most abundant 500 proteins account for 80%. Since it seems likely that the pH 4–8 gels give a representative sampling of all proteins, we estimate that half of the total cellular protein is accounted for by fewer than 100 different gene products, principally glycolytic enzymes and proteins involved in protein synthesis.

**Correlation of protein abundance with mRNA abundance.** Estimates of mRNA abundance for each gene have been made by SAGE (27) and by hybridization of cRNA to oligonucleotide arrays (30). These two methods give broadly similar results; yet each method has strengths and weaknesses (Materials and Methods). Table 1 lists the number of molecules of mRNA per cell for each gene studied. One measurement (mRNA) uses data from SAGE analysis alone (27); a second incorporates data from both SAGE and hybridization (30) (adjusted mRNA) (Table 1; Materials and Methods). We correlated protein abundance with mRNA abundance (Fig. 2). For ad-

justed mRNA versus protein, the Spearman rank correlation coefficient,  $r_s$ , was 0.74 ( $P < 0.0001$ ), and the Pearson correlation coefficient,  $r_p$ , on log transformed data (Materials and Methods) was 0.76 ( $P < 0.00001$ ). We obtained similar correlations for mRNA versus protein and also for other data transformations (Materials and Methods). Thus, several statistical methods show a strong and significant correlation between mRNA abundance and protein abundance. Of course, the correlation is far from perfect; for mRNAs of a given abundance, there is at least a 10-fold range of protein abundance (Fig. 2). Some of this scatter is probably due to posttranscriptional regulation, and some is due to errors in the mRNA or protein data. For example, the protein Yef3 runs poorly on our gels, giving multiple smeared spots. Its abundance has probably been underestimated, partly explaining the low protein/mRNA ratio of Yef3. It is the most extreme outlier in Fig. 2.

These data on mRNA (27, 30) and protein abundance (Table 1) suggest that for each mRNA molecule, there are on average 4,000 molecules of the cognate protein. For instance, for Act1 (actin) there are about 54 molecules of mRNA per cell and about 205,000 molecules of protein. Assuming an mRNA half-life of 30 min (12) and a cell doubling time of 120 min, this suggests that an individual molecule of mRNA might be translated roughly 1,000 times. These calculations are limited to mRNAs for abundant proteins, which are likely to be the mRNAs that are translated best.

A full complement of cell protein is synthesized in about 120 min under these conditions. Thus, 4,000 molecules of protein per molecule of mRNA implies that translation initiates on an mRNA about once every 2 s. This is a remarkably high rate; it implies that if an average mRNA bears 10 ribosomes engaged in translation, then each ribosome completes translation in 20 s; if an average protein has 450 residues; this in turn implies translation of over 20 amino acids per s, a rate considerably higher than estimated for mammals (3 to 8 amino acids per s) (18). These estimates depend on the amount of mRNA per cell (11, 27).

The large number of protein molecules that can be made from a single mRNA raises the issue of how abundance is controlled for less abundant proteins. Many nonabundant proteins may be unstable, and this would reduce the protein/mRNA ratio. In addition, many nonabundant proteins may be translated at suboptimal rates. We have found that mRNAs for nonabundant proteins usually have suboptimal contexts for translational initiation. For example, there are over 600 yeast genes which probably have short open reading frames in the mRNA upstream of the main open reading frame (17a). These may be devices for reducing the amount of protein made from a molecule of mRNA.

**Correlation of codon bias with protein abundance.** The mRNAs for highly expressed proteins preferentially use some codons rather than others specifying the same amino acid (14). This preference is called codon bias. The codons preferred are those for which the tRNAs are present in the greatest amounts. Use of these codons may make translation faster or more efficient and may decrease misincorporation. These effects are most important for the cell for abundant proteins, and so codon bias is most extreme for abundant proteins. The effect can be dramatic—highly biased mRNAs may use only 25 of the 61 codons.

We asked whether the correlation of codon bias with abundance continues for medium-abundance proteins. There are various mathematical expressions quantifying codon bias; here, we have used the CAI (24) (Materials and Methods) because it gives a result between 0 and 1. The  $r_s$  for CAI versus protein abundance is 0.50 ( $P < 0.0001$ ), similar to the mRNA-protein



TABLE 1. Quantitative data\*

Function	Name	CAI	mRNA	Adjusted mRNA	Protein (Glu) (10 <sup>3</sup> )	Protein (Eth) (10 <sup>3</sup> )	E/G ratio
Carbohydrate metabolism	Adh1	0.810	197	197	1,230	972	0.79
	Adh2	0.504	0		0	963	>20
	Cit2	0.185	1	2.8	23	288	12
	Eno1	0.870	No <i>Nla</i>		410	974	2.4
	Eno2	0.892	248	248	650	215	0.33
	Fba1	0.868	179	179	640	608	0.95
	Hxk1,2	0.500	13	10.5	62	46	
	Icl1	0.251	0		0	671	>20
	Pdb1	0.342	5	5	41	33	
	Pdc1	0.903	226	226	280	205	0.73
	Pfk1	0.465	5	5	75	53	0.71
	Pgi1	0.681	14	14	160	120	0.75
	Pyc1	0.260	1	0.7	37	34	
	Tal1	0.579	5	5	110	35	
	Tdh2	0.904	63	63	430	876	NR
	Tdh3	0.924	460	460	1,670	1,927	NR
	Tpi1	0.817	No <i>Nla</i>		No Met	No Met	
Protein synthesis	Efb1	0.762	33	16.5	358	362	
	Eh1,2	0.801	26	26	99	54	0.55
	Prt1	0.303	4	0.7	12	6	
	Rpa0	0.793	246	246	277	100	0.36
	Tif1,2	0.752	29	29	233	106	0.46
	Yef3	0.777	36	36	14	ND	
Heat shock	Hsc82	0.581	2	2.9	112	75	0.67
	Hsp60	0.381	9	2.3	35	82	2.3
	Hsp82	0.517	2	1.3	52	135	2.6
	Hsp104	0.304	7	7	70	161	2.3
	Kar2	0.439	5	10.1	45	102	2.4
	Ssa1	0.709	2	4.3	303	421	1.4
	Ssa2	0.802	10	5	213	324	1.5
	Ssb1,2	0.850	50	50	270	85	
	Ssc1	0.521	2	2.6	68	80	1.2
	Sse1	0.521	8	8	96	48	
	Sti1	0.247	1	1.1	25	44	1.7
Amino acid synthesis	Ade1	0.229	4	4	14	27	
	Ade3	0.276	2	1.7	12	9	
	Ade5,7	0.257	2	1.4	14	4	
	Arg4	0.229	1	8.1	41	41	
	Gdh1	0.585	10	27	148	55	
	Gln1	0.524	11	11	77	104	1.3
	His4	0.267	3	3	15	23	1.5
	Ily5	0.801	6	6	152	109	0.7
	Lys9	0.332	4	4	32	17	0.52
	Met6	0.657	No <i>Nla</i>	22	190	80	0.42
	Pro2	0.248	3	3	30	12	
	Ser1	0.258	2	1.2	15	8	
	Trp5	0.319	5	5	28	12	
Miscellaneous	Act1	0.710	54	54	205	164	0.78
	Adk1	0.531	No <i>Nla</i>		47	43	
	Ald6	0.520	3	3	181	159	
	Atp2	0.424	1	4.1	76	109	1.4
	Bmh1	0.322	46	46	191	137	0.72
	Bmh2	0.384	1	1.4	134	147	
	Cdc48	0.306	2	2.4	32	26	
	Cdc60	0.299	2	0.86	6	2	
	Erg20	0.373	5	5	92	39	
	Gpp1	0.603	16	5	234	158	
	Gsp1	0.621	3	3	115	39	0.34
	Ipp1	0.620	4	4	254	147	0.58
	Lcb1	0.173	0.3	0.8	19	40	
	Mol1	0.423	0	0.45	20	16	
	Pab1	0.488	3	3	41	19	0.47
	Psa1	0.600	15	15	148	56	
	Rnr4	0.497	6	6	44	37	
	Sam1	0.494	5	5	59	21	
	Sam2	0.497	3	15	63	20	
	Sod1	0.376	36	36	631	618	
	Uba1	0.212	2	2	14	20	
	YKL056	0.731	62	62	253	112	0.44
	YLR109	0.549	21	21	930		
	YMR116	0.777	41	41	184	40	0.20

\* CAI, a measure of codon bias, is taken from the YPD mRNA; number of mRNA molecules per cell from SAGE data (27); adjusted mRNA, number of mRNA molecules per cell based on both SAGE and chip hybridization (30) (see Materials and Methods); Protein (Glu), number of molecules of protein per cell in YNB-glucose; Protein (Eth), number of molecules of protein per cell in YNB-ethanol; E/G ratio, ratio of protein abundance in ethanol to glucose. The E/G ratio is not given if it was close to 1 or if it was not repeatable (NR) in multiple gels. Some gene products (e.g., Tif1 and Tif2 [Tif1,2]) were difficult to distinguish on either a protein or an mRNA basis; these are pooled. No *Nla*, there was no suitable *Nla*III site in the 3' region of the gene, and so there are no SAGE mRNA data; No Met, the mature gene product contains no methionines; and so there are no reliable protein data.

TABLE 2. Functions of proteins listed in Table 1

Name <sup>a</sup>	YPD title lines <sup>b</sup>
Adh1	Alcohol dehydrogenase I; cytoplasmic isozyme reducing acetaldehyde to ethanol, regenerating NAD <sup>+</sup>
Adh2	Alcohol dehydrogenase II; oxidizes ethanol to acetaldehyde, glucose repressed
Cit2	Citrate synthase, peroxisomal (nonmitochondrial); converts acetyl-CoA and oxaloacetate to citrate plus CoA
Eno1	Enolase 1 (2-phosphoglycerate dehydratase); converts 2-phospho-D-glycerate to phosphoenolpyruvate in glycolysis
Eno2	Enolase 2 (2-phosphoglycerate dehydratase); converts 2-phospho-D-glycerate to phosphoenolpyruvate in glycolysis
Fba1	Fructose biphosphate aldolase II; sixth step in glycolysis
Hxk1	Hexokinase I; converts hexoses to hexose phosphates in glycolysis; repressed by glucose
Hxk2	Hexokinase II; converts hexoses to hexose phosphates in glycolysis and plays a regulatory role in glucose repression
Icl1	Isocitrate lyase, peroxisomal; carries out part of the glyoxylate cycle; required for gluconeogenesis
Pdb1	Pyruvate dehydrogenase complex, E1 beta subunit
Pdc1	Pyruvate decarboxylase isozyme 1
Pfk1	Phosphofructokinase alpha subunit, part of a complex with Pfk2p which carries out a key regulatory step in glycolysis
Pgi1	Glucose-6-phosphate isomerase, converts glucose-6-phosphate to fructose-6-phosphate
Pyc1	Pyruvate carboxylase 1; converts pyruvate to oxaloacetate for gluconeogenesis
Tal1	Transaldolase; component of nonoxidative part of pentose phosphate pathway
Tdh1	Glyceraldehyde-3-phosphate dehydrogenase 2; converts D-glyceraldehyde 3-phosphate to 1,3-dephosphoglycerate
Tdh3	Glyceraldehyde-3-phosphate dehydrogenase 3; converts D-glyceraldehyde 3-phosphate to 1,3-dephosphoglycerate
Tpi1	Triosephosphate isomerase; interconverts glyceraldehyde-3-phosphate and dihydroxyacetone phosphate
Efb1	Translation elongation factor EF-1β; GTP/GTP exchange factor for Tef1p/Tef2p
Eft1	Translation elongation factor EF-2; contains diphthamide which is not essential for activity; identical to Eft2p
Eft2	Translation elongation factor EF-2; contains diphthamide which is not essential for activity; identical to Eft1p
Pti1	Translation initiation factor eIF3 beta subunit (p90); has an RNA recognition domain
Rpa0 (RPPO)	Acidic ribosomal protein A0
Tif1	Translation initiation factor 4A (eIF4A) of the DEAD box family
Tif2	Translation initiation factor 4A (eIF4A) of the DEAD box family
Yef3	Translation elongation factor EF-3A; member of ATP-binding cassette superfamily
Hsc82	Chaperonin homologous to <i>E. coli</i> HtpG and mammalian HSP90
Hsp60	Mitochondrial chaperonin that cooperates with Hsp10p; homolog of <i>E. coli</i> GroEL
Hsp82	Heat-inducible chaperonin homologous to <i>E. coli</i> HtpG and mammalian HSP90
Hsp104	Heat shock protein required for induced thermotolerance and for resolubilizing aggregates of denatured proteins; important for [psi <sup>+</sup> ]-to-[psi <sup>-</sup> ] prion conversion
Kar2	Heat shock protein of the endoplasmic reticulum lumen required for protein translocation across the endoplasmic reticulum membrane and for nuclear fusion; member of the HSP70 family
Ssa1	Cytoplasmic chaperone; heat shock protein of the HSP70 family
Ssa2	Cytoplasmic chaperone; member of the HSP70 family
Ssb1	Heat shock protein of HSP70 family involved in the translational apparatus
Ssb2	Heat shock protein of HSP70 family, cytoplasmic
Ssc1	Mitochondrial protein that acts as an import motor with Tim44p and plays a chaperonin role in receiving and folding of protein chains during import; heat shock protein of HSP70 family
Sse1	Heat shock protein of the HSP70 family; multicopy suppressor of mutants with hyperactivated Ras/cyclic AMP pathway
Sui1	Stress-induced protein required for optimal growth at high and low temperature; has tetratricopeptide repeats
Ade1	Phosphoribosylamidoimidazole-succinocarboxamide synthase; catalyzes the seventh step in de novo purine biosynthesis pathway
Ade3	C <sub>4</sub> tetrahydrofolate synthase (trifunctional enzyme), cytoplasmic
Ade5,7	Phosphoribosylamine-glycine ligase plus phosphoribosylformylglycinamide cyclo-ligase; bifunctional protein
Arg4	Argininosuccinate lyase; catalyzes the final step in arginine biosynthesis
Gdh1	Glutamate dehydrogenase (NADP <sup>+</sup> ); combines ammonia and α-ketoglutarate to form glutamate
Glu1	Glutamine synthetase; combines ammonia to glutamate in ATP-driven reaction
His4	Phosphoribosyl-AMP cyclohydrolase/phosphoribosyl-ATP pyrophosphohydrolase/histidinol dehydrogenase; 2nd, 3rd, and 10th steps of his biosynthesis pathway
Ilv5	Ketol-acid reductoisomerase (acetophydroxy, acid reductoisomerase) (alpha-keto-β-hydroxylacyl) reductoisomerase; second step in Val and Ile biosynthesis pathway
Lys9	Saccharopine dehydrogenase (NADP <sup>+</sup> , L-glutamate forming) (saccharopine reductase), seventh step in lysine biosynthesis pathway
Met6	Homocysteine methyltransferase; (5-methyltetrahydropteroyl triglutamate-homocysteine methyltransferase), methionine synthase, cobalamin independent
Pro2	γ-Glutamyl phosphate reductase (phosphoglutamate dehydrogenase), proline biosynthetic enzyme
Ser1	Phosphoserine transaminase; involved in synthesis of serine from 3-phosphoglycerate
Tpy5	Tryptophan synthase, last (5th) step in tryptophan biosynthesis pathway
Act1	Actin; involved in cell polarization, endocytosis, and other cytoskeletal functions
Adk1	Adenylate kinase (GTP:AMP phosphotransferase), cytoplasmic
Ald6	Cytosolic acetaldehyde dehydrogenase
Atp2	Beta subunit of F1-ATP synthase; 3 copies are found in each F1 oligomer
Bmh1	Homolog of mammalian 14-3-3 protein; has strong similarity to Bmh2p
Bmh2	Homolog of mammalian 14-3-3 protein; has strong similarity to Bmh1p
Cdc48	Protein of the AAA family of ATPases; required for cell division and homotypic membrane fusion
Cdc60	Leucyl-tRNA synthetase, cytoplasmic
Erg20	Farnesyl pyrophosphate synthetase; may be rate-limiting step in sterol biosynthesis pathway
Gpp1 (Rhr2)	or: Glycerol phosphate phosphatase
Gsp1	Ran, a GTP-binding protein of the Ras superfamily involved in trafficking through nuclear pores
Ipp1	Inorganic pyrophosphatase, cytoplasmic
Lcb1	Component of serine C-palmitoyltransferase; first step in biosynthesis of long-chain base component of sphingolipids
Mot1 (Thi4)	Thiamine-repressed protein essential for growth in the absence of thiamine
Pab1	Poly(A)-binding protein of cytoplasm and nucleus; part of the 3'-end RNA-processing complex (cleavage factor I); has 4 RNA recognition domains
Psa1	Mannose-1-phosphate guanylyltransferase; GDP-mannose pyrophosphorylase
Rnr4	Ribonucleotide reductase small subunit
Sam1	S-Adenosylmethionine synthetase 1
Sam2	S-Adenosylmethionine synthetase 2
Sod1	Copper-zinc superoxide dismutase
Uba1	Ubiquitin-activating (E1) enzyme
YKL056	Resembles translationally controlled tumor protein of animal cells and higher plants
YLR109 (Ahp1)	Alkyl hydroperoxide reductase
YMR116 (Asc1)	Abundant protein with effects on translational efficiency and cell size, has two WD (WD-40) repeats

<sup>a</sup> Accepted name from the *Saccharomyces* genome database and YPD. Names in parentheses represent recent changes.<sup>b</sup> Courtesy of Proteome, Inc., reprinted with permission.

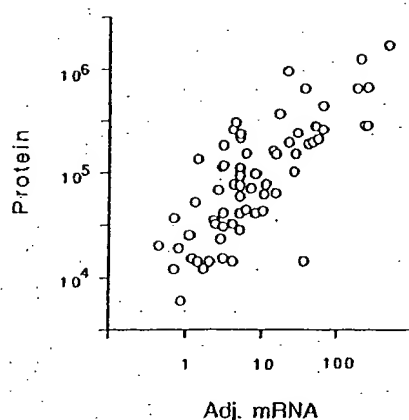


FIG. 2. Correlation of protein abundance with adjusted mRNA abundance. The number of molecules per cell of each protein is plotted against the number of molecules per cell of the cognate mRNA, with an  $r_p$  of 0.76. Note the logarithmic axes. Data for mRNA were taken from references 27 and 30 and combined as described in Materials and Methods.

correlation, confirming a strong correlation between CAI and protein abundance (Fig. 3). The relationship between CAI and protein abundance is log linear from about 1,000,000 to about 10,000 molecules per cell. We have no data for rarer proteins.

It is not clear whether CAI reflects maximum or average levels of protein expression. The proteins used for the CAI-protein correlation included some proteins which were not expressed at maximum levels under the condition of the experiment (Hsc82, Hsp104, Ssa1, Ade1, Arg4, His4, and others). When these proteins were removed from consideration and the correlation between CAI and the remaining (presumably constitutive) proteins was recalculated, the  $r_p$  was essentially unchanged (not shown).

The equation describing the graph in Fig. 3 is  $\log(\text{protein molecules/cell}) = (2.3 \times \text{CAI}) + 3.7$ . Thus, under certain conditions (a CAI of 0.3 or greater; a constitutively expressed gene), a very rough estimate of protein abundance can be made by raising 10 to the power of  $[(2.3 \times \text{CAI}) + 3.7]$ .

The distribution of CAI over the genome (Fig. 4) consists of a lower, bell-shaped distribution, possibly indicating a region where there is no selection for codon bias, and an upper, flat distribution, starting at a CAI of about 0.3, possibly indicating a region where there is selection for codon bias. Almost all of the proteins whose abundance we have measured are in the upper, flat portion of the distribution. In the lower, bell-shaped region, we do not know whether there is a correlation between CAI and protein abundance.

Changes in protein abundance in glucose and ethanol. A comparison of cells grown in glucose (Fig. 1A) with cells grown in ethanol (Fig. 1B) is shown in Table 1. As is well known, some proteins are induced tremendously during growth on ethanol. Two striking examples are the peroxisomal enzymes Icl1 (isocitrate lyase) and Cit2 (citrate synthase), which are induced in ethanol by more than 100- and 12-fold, respectively (Fig. 1; Table 1). These enzymes are key components of the glyoxylate shunt, which diverts some acetyl coenzyme A (acetyl-CoA) from the tricarboxylic acid cycle to gluconeogenesis. *S. cerevisiae* requires large amounts of carbohydrate for its cell wall; in ethanol medium, this carbohydrate comes from gluconeogenesis, which depends on the glyoxylate shunt and on the glycolytic pathway running in reverse. The need for

gluconeogenesis also explains why glycolytic enzymes are abundant even in ethanol medium. Thus, 2D gel analysis shows the prominence of the glycolytic and glyoxylate shunt enzymes in cells grown on ethanol, emphasizing that gluconeogenesis, presumably largely for production of the cell wall, is a major metabolic activity under these conditions.

During gluconeogenesis, substrate-product relationships are reversed for the glycolytic enzymes. One might expect that not all glycolytic enzymes would be well adapted to the reverse reaction. Indeed, 2D gels show that in ethanol, Adh2 (alcohol dehydrogenase 2) is strongly induced (16), while its isozyme Adh1 is not greatly affected. Adh1 and Adh2 each interconvert acetaldehyde and ethanol. Adh1 has a relatively high  $K_m$  for ethanol (17 mM), while Adh2 has a lower  $K_m$  (0.8 mM) (5). Thus, it is thought that Adh1 is specialized for glycolysis (acetaldehyde to ethanol), while Adh2 is specialized for respiration (ethanol to acetaldehyde) (5, 29). Similarly, Eno1 (enolase 1) is induced in ethanol, while its isozyme Eno2 (enolase 2) decreases in abundance (Table 1) (4, 19). Eno1 is inhibited by 2-phosphoglycerate (the glycolytic substrate), while Eno2 is inhibited by phosphoenolpyruvate (the gluconeogenic substrate) (4). Perhaps Eno1 has a lower  $K_m$  for phosphoenolpyruvate than does Eno2, though to our knowledge this has not been tested. Thus, the 2D gels distinguish isozymes specialized for growth on glucose (Adh1 and Eno2) from isozymes specialized for ethanol (Adh2 and Eno1).

Many heat shock proteins (e.g., Hsp60, Hsp82, Hsp104, and Kar2) were about twofold more abundant in ethanol medium than in glucose medium. This is consistent with the increased heat resistance of cells grown in ethanol (3).

Enzymes involved in protein synthesis (Efi1, Rpa0, and Tif1) were about twice as abundant in glucose medium as in ethanol medium. This may reflect the higher growth rate of the cells in glucose.

Phosphorylation of proteins. To examine protein phosphorylation, we labeled cells with  $^{32}\text{P}$  and ran 2D gels to examine phosphoproteins. About 300 distinct spots, probably representing 150 to 200 proteins, could be seen on pH 4–8 gels (Fig. 5B). We then aligned autoradiograms of three gels, each with a different kind of labeled protein ( $^{32}\text{P}$  only [Fig. 5B],  $^{32}\text{P}$  plus  $^{35}\text{S}$  [Fig. 5A], and  $^{35}\text{S}$  only [not shown, but see Fig. 1 for example]). In this way, we made provisional identification of

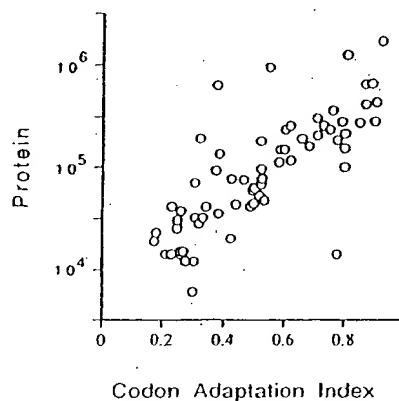


FIG. 3. Correlation of protein abundance with CAI. The number of molecules per cell of each protein is plotted against the CAI for that protein. Note the logarithmic scale on the protein axis. Data for the CAI are from the YPD database (13).

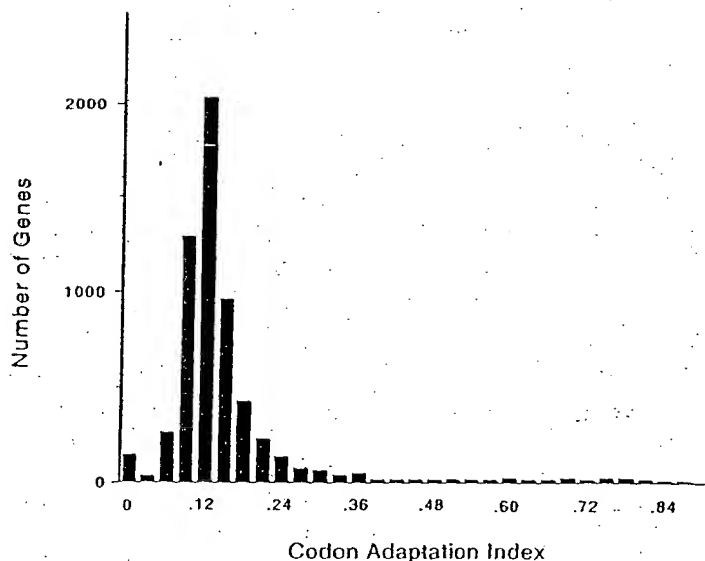


FIG. 4. Distribution of CAI over the whole genome, shown in intervals of 0.030 (i.e., there are 150 genes with a CAI between 0.000 and 0.030, inclusive; 31 genes with a CAI between 0.031 and 0.060; 269 genes with a CAI between 0.061 and 0.090; 1,296 genes with a CAI between 0.091 and 0.120, etc.). The distribution peaks with 2,028 genes with a CAI between 0.121 and 0.150.

some of the  $^{32}\text{P}$ -labeled spots as particular  $^{35}\text{S}$ -labeled spots. All such identifications are somewhat uncertain, since precise alignments are difficult, and of course multiple spots may exactly comigrate. Nevertheless, we believe that most of the provisional identifications are probably correct. Among the major  $^{32}\text{P}$ -labeled proteins are the hexokinases Hxk1 and Hxk2, the acidic ribosome-associated protein Rpa0, the translation factors Yef3 and Efb1, and probably Hsp70 heat shock proteins of the Ssa and Ssb families. Rpa0 and Efb1 are quantitatively monophosphorylated.

Many yeast proteins resolve into multiple spots on these 2D gels (7). Yef3 has five or more spots, at least four of which comigrate with  $^{32}\text{P}$ . Tif1 has a major spot showing no  $^{32}\text{P}$  labeling and a minor, more acidic spot which overlaps with some  $^{32}\text{P}$  label. Tif1 has at least seven spots (7); two of these overlap with some  $^{32}\text{P}$  label, but five do not (Fig. 5). Efb1 has at least three spots (7), and none of these overlap with  $^{32}\text{P}$ , although there are three nearby, unidentified  $^{32}\text{P}$ -labeled spots (a, c, and d in Fig. 5). Spots that seem to be extra forms of Met6, Pdc1, Eno2, and Fba1 can be seen in Fig. 6A, but there is little  $^{32}\text{P}$  at these positions in Fig. 5. Thus, phosphorylation explains some but not all of the different protein isoforms seen.

The cell cycle is regulated in part by phosphorylation. We compared  $^{32}\text{P}$ -labeled proteins from cells synchronized in  $G_1$  with  $\alpha$ -factor, in cells synchronized in  $G_1$  by depletion of  $G_1$  cyclins, and in cells synchronized in M phase with nocodazole. Only very minor differences were seen, and these were difficult to reproduce. The cell cycle proteins regulated by phosphorylation may not be abundant enough for this technique to be applied easily.

**Centrifugal fractionation.** We fractionated  $^{35}\text{S}$ -labeled extracts by centrifugation (Materials and Methods). Figure 6A shows the proteins in the supernatant of a high-speed ( $100,000 \times g$ , 30 min) centrifugation, while Fig. 6B shows the proteins in the pellet of a low-speed ( $16,000 \times g$ , 10 min) centrifugation. Many proteins are tremendously enriched in one fraction or the other, while others are present in both.

Most glycolytic enzymes (e.g., Tdh2, Tdh3, Eno2, Pdc1, Adh1, and Fba1) are enriched in the supernatant fraction. The only exception is Pfk1 (not indicated), which is found in both pellet and supernatant fractions. Many proteins involved in protein synthesis (Eft1, Yef3, Pti1, Tif1, and Rpa0) are in the pellet, possibly because of the association of ribosomes with the endoplasmic reticulum. However, Efb1 is in the supernatant, as is a substantial portion of the Eft1. Perhaps surprisingly, several mitochondrial proteins (Atp2 [not shown] and Ilv5) are largely in the supernatant. Perhaps glass bead breakage of cells releases mitochondrial proteins. The nuclear protein Gsp1 is in the pellet fraction. The enrichment produced by centrifugation makes it possible to see minor spots which are otherwise poorly resolved from surrounding proteins. Figure 6B shows that the previously identified Tif1 spot is surrounded by as many as six other spots that cofractionate. We observed six identical or very similar additional spots when we overexpressed Tif1 from a high-copy-number plasmid (not shown). Signal overlaps only one or two of these spots in  $^{32}\text{P}$ -labeling experiments (Fig. 5), and so the different forms are not mainly due to different phosphorylation states.

## DISCUSSION

Our experience with developing a 2D gel protein database for *S. cerevisiae* is summarized here. With current technology, we can see the most abundant 1,200 proteins, which is about one-third to one-quarter of the proteins expressed. The remaining proteins will be difficult to see and study with the methods that we have used, not because of a lack of sensitivity but because weak spots are covered by nearby strong spots.

Of the 1,200 proteins seen, we have identified 148, with a bias toward the most abundant proteins. Steady application of the methods already used would allow identification of most of the remaining proteins. Gene overexpression will be particularly useful, since it is not affected by the lower abundance of the remaining visible proteins.

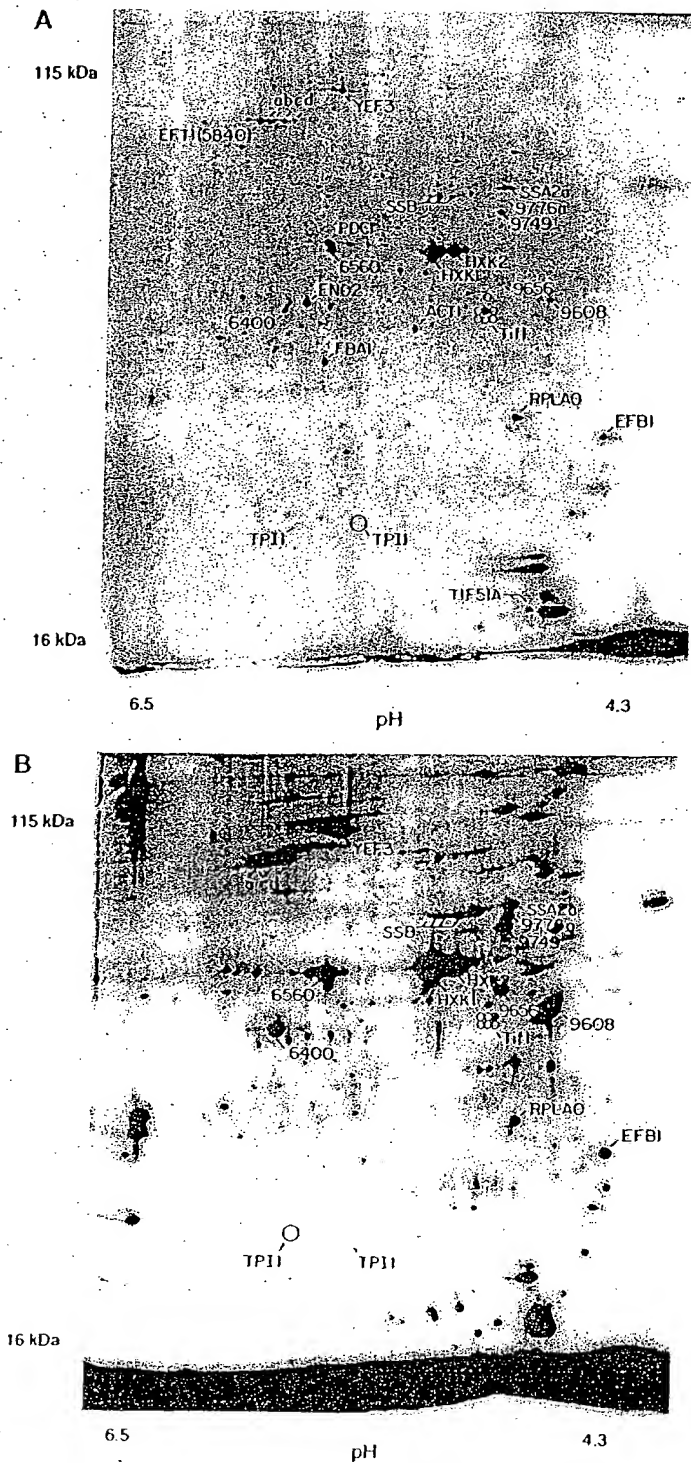


FIG. 5. Phosphorylated proteins. (A) Mixture of  $^{32}\text{P}$ -labeled proteins and  $^{35}\text{S}$ -labeled proteins. Two separate labeling reactions were done, one with  $^{32}\text{P}$  and one with  $^{35}\text{S}$ , and extracts were mixed and run on a 2D gel. Spots marked with numbers rather than gene names represent spots noted on  $^{35}\text{S}$  gels but unidentified. Spots labeling with  $^{32}\text{P}$  were identified by (i) increased labeling compared to the  $^{35}\text{S}$ -only gel (not shown); (ii) the characteristic fuzziness of a  $^{32}\text{P}$ -labeled spot; and (iii) the decay of signal intensity seen on exposures made 4 weeks later (not shown). A minor form of Tpi1 and at least six minor forms of Tif1 have been noted in overexpression experiments (see also Fig. 6B); positions of the minor forms are indicated by circles. (B)  $^{32}\text{P}$ -only labeling. The major form of Tpi1, which is not labeled with  $^{32}\text{P}$ , is indicated by a large circle; positions of seven forms of Tif1 are indicated by smaller circles.

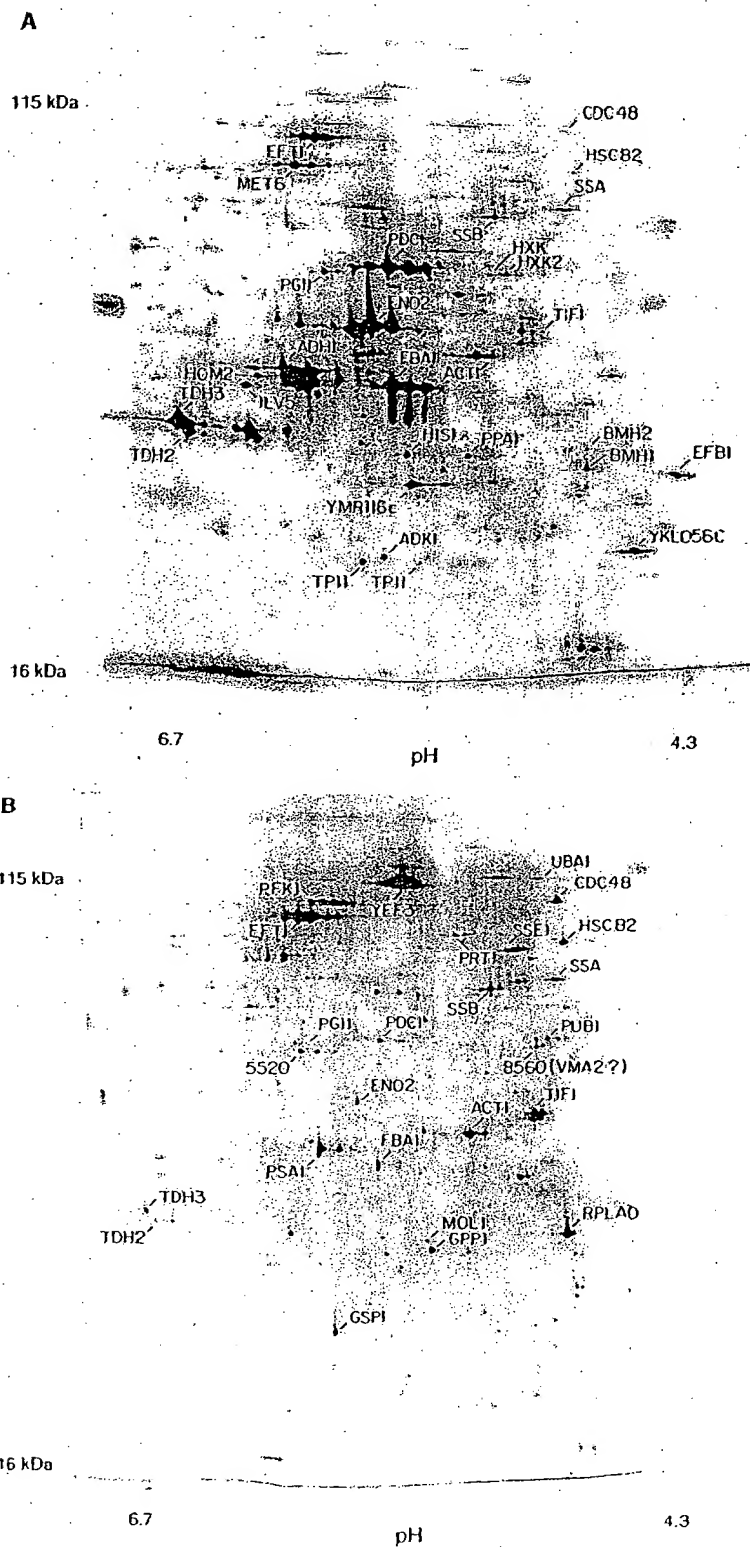


FIG. 6. Fractionation by centrifugation. (A) Proteins in the supernatant of a  $100,000 \times g$ , 30-min spin; proteins in the pellet of a  $16,000 \times g$ , 10-min spin. Supernatant fractions examined in multiple experiments done over a wide range of  $g$  forces looked similar to each other, as did the pellet fractions.

2D gels of the kind that we have used are not suitable for visualization of rare proteins. However it will be possible to study on a global basis metabolic processes involving relatively abundant proteins, such as protein synthesis, glycolysis, gluconeogenesis, amino acid synthesis, cell wall synthesis, nucleotide synthesis, lipid metabolism, and the heat shock response.

Gygi et al. (10) have recently completed a study similar to ours. Despite generating broadly similar data, Gygi et al. reached markedly different conclusions. We believe that both mRNA abundance and codon bias are useful predictors of protein abundance. However, Gygi et al. feel that mRNA abundance is a poor predictor of protein abundance and that "codon bias is not a predictor of either protein or mRNA levels" (10). These different conclusions are partly a matter of viewpoint. Gygi et al. focus on the fact that the correlations of mRNA and codon bias with protein abundance are far from perfect, while we focus on the fact that, considering the wide range of mRNA and protein abundance and the undoubted presence of other mechanisms affecting protein abundance, the correlations are quite good.

However, the different conclusions are also partly due to different methods of statistical analysis and to real differences in data. With respect to statistics, Gygi et al. used the Pearson product-moment correlation coefficient ( $r_p$ ) to measure the covariance of mRNA and protein abundance. Depending on the subset of data included, their  $r_p$  values ranged from 0.1 to 0.94. Because of the low  $r_p$  values with some subsets of the data, Gygi et al. concluded that the correlation of mRNA to protein was poor. However, the  $r_p$  correlation is a parametric statistic and so requires variates following a bivariate normal distribution; that is, it would be valid only if both mRNA and protein abundances were normally distributed. In fact, both distributions are very far from normal (data not shown), and so a calculation of  $r_p$  is inappropriate. There was no statistical backing for the assertion that codon bias fails to predict protein abundance.

We have taken two statistical approaches. First, we have used the Spearman rank correlation coefficient ( $r_s$ ). Since this statistic is nonparametric, there is no requirement for the data to be normally distributed. Using the  $r_s$ , we find that mRNA abundance is well correlated with protein abundance ( $r_s = 0.74$ ), and the CAI is also well correlated with protein abundance ( $r_s = 0.80$ ) (and also with mRNA abundance [data not shown]). For the data of Gygi et al. (10), we obtained similar results, though with their data the correlation is not as good;  $r_s = 0.59$  for the mRNA-to-protein correlation, and  $r_s = 0.59$  for the codon bias-to-protein correlation.

In a second approach, we transformed the mRNA and protein data to forms where they were normally distributed, to allow calculation of an  $r_p$  (Materials and Methods). Two transformations, Box-Cox and logarithmic, were used; both gave good correlations with our data [e.g.,  $r_p = 0.76$  for  $\log(\text{adjusted RNA})$  to  $\log(\text{protein})$ ]. We were not able to transform the data of Gygi et al. to a normal distribution.

Finally, there are also some differences in data between the two studies. These may be partly due to the different measurement techniques used: Gygi et al. measured protein abundance by cutting spots out of gels and measuring the radioactivity in each spot by scintillation counting, whereas we used phosphorimaging of intact gels coupled to image analysis. We compared our data to theirs for the proteins common between the studies (but excluding proteins whose mRNAs are known to differ between rich and minimal media, and excluding Tif1, which was anomalous in differing by 100-fold between the two data sets). The  $r_s$  between the two protein data sets was 0.88 ( $P < 0.0001$ ). Although this is a strong correlation, the fact that

it is less than 1.0 suggests that there may have been errors in measuring protein abundance in one or both studies. After normalizing the two data sets to assume the same amount of protein per cell, we found a systematic tendency for the protein abundance data of Gygi et al. to be slightly higher than ours for the highest-abundance proteins and also for the lowest-abundance proteins but slightly lower than ours for the middle-abundance proteins. These systematic differences suggest some systematic errors in protein measurement. Although we do not know what the errors are, we suggest the following as a reasonable speculation. For the highest-abundance proteins, we may have underestimated the amount of protein because of a slightly nonlinear response of the phosphorimager screens. For the lowest-abundance proteins, Gygi et al. may have overestimated the amount of protein because of difficulties in accurately cutting very small spots out of the gel and because of difficulties in background subtraction for these small, weak spots. The difference in the middle abundance proteins may be a consequence of normalization, given the two errors above.

The low-abundance proteins in the data set of Gygi et al. have a poor correlation with mRNA abundance. We calculate that the  $r_s$  is 0.74 for the top 54 proteins of Gygi et al. but only 0.22 for the bottom 53 proteins, a statistically significant difference. However, with our data set, the  $r_s$  is 0.62 for the top 33 proteins and 0.56 (not significantly different) for the bottom 33 proteins (which are comparable in abundance to the bottom 53 proteins of Gygi et al.). Thus, our data set maintains a good correlation between mRNA and protein abundance even at low protein abundance. This is consistent with our speculation that protein quantification by phosphorimaging and image analysis may be more accurate for small, weak spots than is cutting out spots followed by scintillation counting. Our relatively good correlations even for nonabundant proteins may also reflect the fact that we used both SAGE data and RNA hybridization data, which is most helpful for the least abundant mRNAs. In summary, we feel that the poor correlation of protein to mRNA for the nonabundant proteins of Gygi et al. may reflect difficulty in accurately measuring these nonabundant proteins and mRNAs, rather than indicating a truly poor correlation *in vivo*. It is not surprising that observed correlations would be poorer with less-abundant proteins and mRNAs, simply because the accuracy of measurement would be worse.

How well can mRNA abundance predict protein abundance? With  $r_p = 0.76$  for logarithmically transformed mRNA and protein data, the coefficient of determination,  $(r_p)^2$ , is 0.58. This means that more than half (in log space) of the variation in protein abundance is explained by variation in mRNA abundance. When converted back to arithmetic values, protein abundances vary over about 200-fold (Table 1), and  $(r_p)^2 = 0.58$  for the log data means that of this 200-fold variation, about 20-fold is explained by variation in the abundance of mRNA and about 10-fold is unexplained (but could be due partly to measurement errors). For proteins much less abundant than those considered here, we imagine the *in vivo* correlation between mRNA and protein abundance will be worse, and other regulatory mechanisms such as protein turnover will be more important.

Some important conclusions can be drawn from this sampling of the proteome. First, there is an enormous range of protein abundance, from nearly 2,000,000 molecules per cell for some glycolytic enzymes to about 100 per cell for some cell cycle proteins (26a). Second, about half of all cellular protein is found in fewer than 100 different gene products, which are mostly involved in carbohydrate metabolism or protein synthe-

sis. Third, the correlation between protein abundance and CAI is log linear as far as we can see, which is from about 10,000 protein molecules per cell to about 1,000,000. This is somewhat surprising, because it implies that selective forces for codon bias are significant even at moderate expression levels. It also means that codon bias is a useful predictor of protein abundance even for moderately low bias proteins. Fourth, there is a good correlation between protein abundance and mRNA abundance for the proteins that we have studied. This validates the use of mRNA abundance as a rough predictor of protein abundance, at least for relatively abundant proteins. Fifth, for these abundant proteins, there are about 4,000 molecules of protein for each molecule of mRNA. This last conclusion raises questions as to how the levels of nonabundant proteins are regulated and suggests that protein instability, regulated translation, suboptimal rates of translation, and other mechanisms in addition to transcriptional control may be very important for these proteins.

#### ACKNOWLEDGMENTS

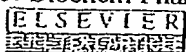
We thank Neena Sareen and Nick Bizios (CSHL 2D gel laboratory) for production of 2D gels, Tom Volpe for help with some experiments, Corine Driessens for help with calculations and statistics, and Herman Wijnen and Nick Edgington for comments on the manuscript. We especially thank Tim Tully for in-depth statistical analysis and for insightful discussions on statistical interpretations.

This work was supported by grant P41-RR02188 from the NIH Biomedical Research Technology Program, Division of Research Resources, to J.I.G., by Small Business Innovation Research grant R44 GM54110 to Proteome, Inc., by grant DAMD17-94-J4050 from the Army Breast Cancer Program to B.F., and by NIH grant RO1 GM45410 to B.F.

#### REFERENCES

- Baroni, M. D., E. Martegani, P. Monti, and L. Alberghina. 1989. Cell size modulation by *CDC25* and *R452* genes in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 9:2715-2723.
- Boucherie, H., F. Sogliocco, R. Joubert, J. Maillet, J. Labarre, and M. Perrot. 1996. Two-dimensional gel protein database of *Saccharomyces cerevisiae*. *Electrophoresis* 17:1683-1699.
- Elliott, B., and B. Futcher. 1993. Stress resistance of yeast cells is largely independent of cell cycle phase. *Yeast* 9:33-42.
- Enfian, K. D., B. Meurer, H. Kohler, K. H. Mann, and D. Mecke. 1987. Studies on the regulation of enolases and compartmentation of cytosolic enzymes in *Saccharomyces cerevisiae*. *Biochim. Biophys. Acta* 923:214-221.
- Ganzhorn, A. J., D. W. Green, A. D. Hershey, R. M. Gould, and B. V. Plapp. 1987. Kinetic characterization of yeast alcohol dehydrogenases. Amino acid residue 294 and substrate specificity. *J. Biol. Chem.* 262:3754-3761.
- Garrels, J. I. 1989. The Quest system for quantitative analysis of two-dimensional gels. *J. Biol. Chem.* 264:5269-5282.
- Garrels, J. I., B. Futcher, R. Kobayashi, G. I. Latter, B. Schwender, T. Volpe, J. R. Warner, and C. S. McLaughlin. 1994. Protein identifications for a *Saccharomyces cerevisiae* protein database. *Electrophoresis* 15:1466-1486.
- Garrels, J. I., C. S. McLaughlin, J. R. Warner, B. Futcher, G. I. Latter, R. Kobayashi, B. Schwender, T. Volpe, D. S. Anderson, R. Mesquita-Fuentes, and W. E. Payne. 1997. Proteome studies of *S. cerevisiae*: identification and characterization of abundant proteins. *Electrophoresis* 18:1347-1360.
- Goffeau, A., B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Holtzman, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, and S. G. Oliver. 1996. Life with 6000 genes. *Science* 274:563-567.
- Gy6, S. P., Y. Rochon, B. R. Franza, and R. Aebersold. 1999. Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* 19:1720-1730.
- Hereford, L. M., and M. Rosbash. 1977. Number and distribution of polyadenylated RNA sequences in yeast. *Cell* 10:453-462.
- Herrick, D., R. Parker, and A. Jacobson. 1990. Identification and comparison of stable and unstable mRNAs in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 10:2269-2284.
- Hodges, P. E., A. H. McKee, B. P. Davis, W. E. Payne, and J. I. Garrels. 1999. The Yeast Proteome Database (YPD): a model for the organization of genome-wide functional data. *Nucleic Acids Res.* 27:69-73.
- Ikemura, T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.* 2:13-34.
- Johnston, G. C., F. R. Pringle, and L. H. Hartwell. 1977. Coordination of growth with cell division in the yeast *S. cerevisiae*. *Exp. Cell Res.* 105:79-98.
- Johnston, M., and M. Carlson. 1992. Regulation of carbon and phosphate utilization, p. 193-281. In E. Jones, J. Pringle, and J. Broach (ed.), *The molecular and cellular biology of the yeast Saccharomyces*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Kornblatt, M. J., and A. Klugerman. 1989. Characterization of the enolase isozymes of rabbit brain: kinetic differences between mammalian and yeast enolases. *Biochem. Cell. Biol.* 67:103-107.
- Latter, G., and B. Futcher. Unpublished data.
- Mathews, B., N. Sonenberg, and J. W. B. Hershey. 1996. Origins and targets of translational control, p. 1-29. In J. W. B. Hershey, M. B. Mathews, and N. Sonenberg (ed.), *Translational control*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- McAlister, L., and M. J. Holland. 1982. Targeted deletion of a yeast enolase structural gene. Identification and isolation of yeast enolase isozymes. *J. Biol. Chem.* 257:7181-7188.
- Monardo, P. J., T. Boutell, J. I. Garrels, and G. I. Latter. 1994. A distributed system for two-dimensional gel analysis. *Comput. Appl. Biosci.* 10:137-143.
- O'Farrell, P. H. 1975. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* 250:4007-4021.
- Patterson, S. D., and G. I. Latter. 1993. Evaluation of storage phosphor imaging for quantitative analysis of 2-D gels using the Quest II system. *BioTechniques* 15:1076-1083.
- Sagliocco, F., J. C. Guillemot, C. Monribot, J. Capdevielle, M. Perrot, E. Ferran, P. Ferrara, and H. Boucherie. 1996. Identification of proteins of the yeast protein map using genetically manipulated strains and peptide-mass fingerprinting. *Yeast* 12:1519-1533.
- Sharp, P. M., and W. H. Li. 1987. The Codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 15:281-295.
- Shevchenko, A., O. N. Jensen, A. V. Poulsen, F. Sogliocco, M. Wilm, O. Vorm, P. Mortensen, A. Shevchenko, H. Boucherie, and M. Mann. 1996. Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc. Natl. Acad. Sci. USA* 93:14440-14445.
- Thomas, B. J., and R. Rothstein. 1989. Elevated recombination rates in transcriptionally active DNA. *Cell* 56:619-630.
- Tyers, M., and B. Futcher. Unpublished data.
- Velculescu, V. E., L. Zhong, W. Zhou, J. Vogelstein, M. A. Basrai, D. E. Bassett, Jr., P. Hieter, B. Vogelstein, and K. W. Kinzler. 1997. Characterization of the yeast transcriptome. *Cell* 88:243-251.
- Warner, J. 1991. Labeling of RNA and phosphoproteins in *S. cerevisiae*. *Methods Enzymol.* 194:423-428.
- Wills, C. 1976. Production of yeast alcohol dehydrogenase isoenzymes by selection. *Nature* 261:26-29.
- Wodicka, L. Personal communication.
- Wodicka, L. Unpublished data.
- Wodicka, L., H. Dong, M. Mittmann, M.-H. Jio, and D. J. Lockhart. 1997. Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nat. Biotechnol.* 15:1359-1367.





**Pre-translational regulation of cytochrome P450 genes is responsible for disease-specific changes of individual P450 enzymes among patients with cirrhosis.**

**George J. Liddle C, Murray M, Byth K, Farrell GC.**

Department of Gastroenterology and Hepatology, University of Sydney at Westmead Hospital, NSW, Australia.

We have recently reported that disease-specific differential alterations in the hepatic expression of xenobiotic-metabolizing cytochrome P450 (CYP P450) enzymes occur in patients with advanced liver disease. In order to determine whether the observed changes in CYP proteins are modulated at pre- or post-translational levels, we have now examined the hepatic levels of mRNA for CYPs 1A2, 2C9, 2E1 and 3A4 by solution hybridization in the same livers of 20 controls (surgical waste from histologically normal livers), 32 cases of hepatocellular and 18 of cholestatic severe chronic liver disease. CYP1A2 mRNA and CYP1A immunoreactive protein were both reduced in livers with hepatocellular and cholestatic types of cirrhosis. In contrast, CYP3A4 mRNA and protein were reduced only in livers from patients with hepatocellular diseases. For 1A2 and 3A4 there were significant correlations between mRNA species and the respective protein contents ( $r_{S1A2} = 0.74$ ,  $r_{S3A4} = 0.64$ ,  $P < 0.0001$ ). CYP2C9 mRNA was reduced in patients with both cholestatic and hepatocellular types of liver disease, but 2C protein was reduced only in patients with cholestatic dysfunction. The correlation between CYP2C9 mRNA and protein, was also significant ( $r_s = 0.36$ ,  $P < 0.005$ ) but mRNA levels accounted for only 13% of the variability in protein rankings. This is probably a consequence of other CYP2C proteins apart from 2C9 being detected by the anti-2C antibody. CYP2E1 mRNA and protein were reduced in patients with cholestatic liver disease, but in hepatocellular disease the expression of only CYP2E1 mRNA was decreased. CYP2E1 mRNA was significantly correlated with CYP2E1 protein but accounted for only 18% of the variability in protein rankings ( $r_s = 0.43$ ,  $P < 0.0005$ ). Taken collectively these data indicate that the disease-specific alterations of xenobiotic-metabolizing CYP enzymes among patients with cirrhosis is due, at least in part, to pre-translational mechanisms. The lack of a strong correlation between CYP2E1 mRNA and protein suggests that this gene, like its rat orthologue, may be subject to pre-translational as well as translational and/or post-translational regulation.

PMID: 7741759 [PubMed - indexed for MEDLINE]

Full text article at  
[www.jimmunol.org](http://www.jimmunol.org)

### **Cyclooxygenase-2 expression in macrophages: modulation by protein kinase C- $\alpha$ .**

**Giroux M, Descoteaux A.**

Institut National de la Recherche Scientifique-Institut Armand-Frappier, Université du Québec, Laval, Canada.

Cyclooxygenase-2 (COX-2) is an inducible enzyme responsible for high levels of PG production during inflammation and immune responses. Previous studies with pharmacological inhibitors suggested a role for protein kinase C (PKC) in PG production possibly by regulating COX-2 expression. In this study, we addressed the role of PKC- $\alpha$  in the modulation of COX-2 expression and PGE<sub>2</sub> synthesis by the overexpressing of a dominant-negative (DN) mutant of this isoenzyme in the mouse macrophage cell line RAW 264.7. We investigated the effect of various stimuli on COX-2 expression, namely, LPS, IFN- $\gamma$ , and the intracellular parasite *Leishmania donovani*. Whereas LPS-induced COX-2 mRNA and protein expression were down-regulated in DN PKC- $\alpha$ -overexpressing clones, IFN- $\gamma$ -induced COX-2 expression was up-regulated in DN PKC- $\alpha$ -overexpressing clones with respect to normal RAW 264.7 cells. Measurements of PGE<sub>2</sub> levels revealed a strong correlation between PGE<sub>2</sub> secretion and IFN- $\gamma$ -induced COX-2 mRNA and protein levels in DN PKC- $\alpha$ -overexpressing clones. Taken together, these results suggest a role for PKC- $\alpha$  in the modulation of LPS- and IFN- $\gamma$ -induced COX-2 expression, as well as in IFN- $\gamma$ -induced PGE<sub>2</sub> secretion.

PMID: 11034408 [PubMed - indexed for MEDLINE]

Comment in:

- [Blood. 2003 Aug 15;102\(4\):1550-1.](#)



**Transcript profiling of human platelets using microarray and serial analysis of gene expression.**

Gnatenko DV, Dunn JJ, McCorkle SR, Weissmann D, Perrotta PL, Bahou WF.

Department of Medicine, Program in Genetics, State University of New York, Stony Brook 11794-8151, USA.

Human platelets are anucleate blood cells that retain cytoplasmic mRNA and maintain functionally intact protein translational capabilities. We have adapted complementary techniques of microarray and serial analysis of gene expression (SAGE) for genetic profiling of highly purified human blood platelets. Microarray analysis using the Affymetrix HG-U95Av2 approximately 12 600-probe set maximally identified the expression of 2147 (range, 13%-17%) platelet-expressed transcripts, with approximately 22% collectively involved in metabolism and receptor/signaling, and an overrepresentation of genes with unassigned function (32%). In contrast, a modified SAGE protocol using the Type IIS restriction enzyme MmeI (generating 21-base pair [bp] or 22-bp tags) demonstrated that 89% of tags represented mitochondrial (mt) transcripts (enriched in 16S and 12S ribosomal RNAs), presumably related to persistent mt-transcription in the absence of nuclear-derived transcripts. The frequency of non-mt SAGE tags paralleled average difference values (relative expression) for the most "abundant" transcripts as determined by microarray analysis, establishing the concordance of both techniques for platelet profiling. Quantitative reverse transcription-polymerase chain reaction (PCR) confirmed the highest frequency of mt-derived transcripts, along with the mRNAs for neurogranin (NGN, a protein kinase C substrate) and the complement lysis inhibitor clusterin among the top 5 most abundant transcripts. For confirmatory characterization, immunoblots and flow cytometric analyses were performed, establishing abundant cell-surface expression of clusterin and intracellular expression of NGN. These observations demonstrate a strong correlation between high transcript abundance and protein expression, and they establish the validity of transcript analysis as a tool for identifying novel platelet proteins that may regulate normal and pathologic platelet (and/or megakaryocyte) functions.

PMID: 12433680 [PubMed - indexed for MEDLINE]

## Transcript profiling of human platelets using microarray and serial analysis of gene expression

Dmitri V. Gnatlenko, John J. Dunn, Sean R. McCorkle, David Weissmann, Peter L. Perrotta, and Wadie F. Bahou

Human platelets are anucleate blood cells that retain cytoplasmic mRNA and maintain functionally intact protein translational capabilities. We have adapted complementary techniques of microarray and serial analysis of gene expression (SAGE) for genetic profiling of highly purified human blood platelets. Microarray analysis using the Affymetrix HG-U95Av2 approximately 12 600-probe set maximally identified the expression of 2147 (range, 13%-17%) platelet-expressed transcripts, with approximately 22% collectively involved in metabolism and receptor/signaling, and an overrepresentation of genes with unassigned function (32%). In contrast, a modified SAGE protocol using the Type IIS restriction enzyme

MmeI (generating 21-base pair [bp] or 22-bp tags) demonstrated that 89% of tags represented mitochondrial (mt) transcripts (enriched in 16S and 12S ribosomal RNAs), presumably related to persistent mt-transcription in the absence of nuclear-derived transcripts. The frequency of non-mt SAGE tags paralleled average difference values (relative expression) for the most "abundant" transcripts as determined by microarray analysis, establishing the concordance of both techniques for platelet profiling. Quantitative reverse transcription-polymerase chain reaction (PCR) confirmed the high frequency of mt-derived transcripts, along with the mRNAs for neurogranin (NGN, a protein kinase C substrate) and

the complement lysis inhibitor clusterin among the top 5 most abundant transcripts. For confirmatory characterization, immunoblots and flow cytometric analyses were performed, establishing abundant cell-surface expression of clusterin and intracellular expression of NGN. These observations demonstrate a strong correlation between high transcript abundance and protein expression, and they establish the validity of transcript analysis as a tool for identifying novel platelet proteins that may regulate normal and pathologic platelet (and/or megakaryocyte) functions. (Blood. 2003;101:2285-2293)

© 2003 by The American Society of Hematology

### Introduction

Human blood platelets play critical roles in normal hemostatic processes and pathologic conditions such as thrombosis, vascular remodeling, inflammation, and wound repair. Generated as cytoplasmic buds from precursor bone marrow megakaryocytes, platelets are anucleate and lack nuclear DNA, although they retain megakaryocyte-derived mRNAs.<sup>1,2</sup> Platelets contain rough endoplasmic reticulum and polyribosomes, and they retain the ability for protein biosynthesis from cytoplasmic mRNA.<sup>3</sup> Quiescent platelets generally display minimal translational activity, although newly formed platelets such as those found in patients with immune thrombocytopenic purpura (ITP) synthesize various  $\alpha$ -granule and membrane glycoproteins (GPs), including GPIb and GPIIb/IIIa ( $\alpha_{IIb}\beta_3$ ). Furthermore, stimulation of quiescent platelets by agonists such as  $\alpha$ -thrombin increases protein synthesis of various platelet proteins, including Bcl-3.<sup>4</sup> Like nucleated cells, the rapid translation of preexisting mRNAs may be regulated by integrin ligation to extracellular matrices.<sup>5</sup> In the case of platelets, the primary integrin involved in this process appears to be  $\alpha_{IIb}\beta_3$  with cooperative signals mediated by the collagen receptor  $\alpha_2\beta_1$ .<sup>6,7</sup>

Integrin-mediated platelet protein synthesis appears to be regulated at the level of translation initiation involving the eukaryotic initiation factor 4E (eIF4E). Instead of directly influencing eIF4E activity via posttranslational modifications (ie, phosphorylation), platelet eIF4E activity best correlates with its spatial redistribution to the mRNA-enriched cytoskeleton.<sup>8</sup> Furthermore, because protein translation is partially inhibited by the immunosuppressant rapamycin, it suggests that adhesion- and/or aggregation-induced outside-in-signaling function to regulate protein synthesis through the mTOR (mammalian target of rapamycin) pathway.<sup>8,9</sup>

Despite the biologic importance of platelets and their intact protein synthetic capabilities, remarkably little is known about platelet mRNAs. Younger platelets contain larger amounts of mRNA with a greater capacity for protein synthesis, as determined by using fluorescent nucleic acid dyes such as thiazole orange.<sup>10</sup> This assay has been used as a quantitative determinant of younger or "reticulated" platelets (RPs). Indeed increased reticulated platelets are typically found in patients with conditions associated with rapid platelet turnover such as ITP; typically RP percentages in

From the Department of Medicine, Department of Pathology, and Program in Genetics, State University of New York, Stony Brook; Biology Department, Brookhaven National Laboratory, Upton, NY; and Department of Pathology, Robert Wood Johnson Medical Center, New Brunswick, NJ.

Submitted September 16, 2002; accepted November 3, 2002. Prepublished online as Blood First Edition Paper, November 14, 2002; DOI 10.1182/blood-2002-09-2797.

Supported by grants HL49141 and HL53665, by a Veteran's Administration REAP award (W.F.B.), and by National Institutes of Health Center grant MO1 10710-5 to the University Hospital General Clinical Research Center. W.B. is an Established Investigator of the American Heart Association. Studies at

Brookhaven National Laboratory were supported by a Laboratory Directed Research and Development award (J.J.D.) and by the Offices of Biological and Environmental Research, and of Basic Energy Sciences (Division of Energy Biosciences) of the US Department of Energy.

Reprints: Wadie F. Bahou, Division of Hematology, HSCT15-040, State University of New York at Stony Brook, Stony Brook, NY 11794-8151; e-mail: wbahou@notes.cc.sunysb.edu.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 U.S.C. section 1734.

© 2003 by The American Society of Hematology

such patients approach 10% to 20% of all platelets, considerably higher than in healthy control subjects.<sup>11</sup> Interestingly, high RPs have been associated with enhanced thrombotic risk when identified in patients with thrombocytosis,<sup>10</sup> suggesting that quantitatively increased mRNA levels may be associated with the prothrombotic phenotype. Whether this is related to globally altered gene expression profiles or to select changes more evident during situations of rapid platelet turnover remains unknown. Certainly, technical limitations of this assay limit its utility in defining prothrombotic genotypes,<sup>10-12</sup> and it cannot identify differentially expressed genes that may be causally implicated in disordered platelet phenotypes.

Toward the goal of defining the molecular anatomy of the platelet genome, we have adapted complementary techniques of microarray and serial analysis of gene expression (SAGE) for genetic profiling of highly purified human blood platelets. Microarray technology represents a "closed" profiling strategy limited by the target genes imprinted onto gene chips. In contrast, SAGE is an "open" architectural system that can be used to identify novel genes and to quantify differentially expressed mRNAs.<sup>13-15</sup> The sequence of each tag along with its positional location uniquely identifies the gene from which it is derived, and differentially expressed genes can be identified in a quantitative manner because the tag frequency reflects the mRNA level at the time of cellular harvest and analysis. By using both technologies, we have identified a number of previously uncharacterized genes that appear to be expressed in human platelets, while simultaneously establishing the dominant frequency of mitochondrial-expressed genomes comprising the platelet mRNA pool. These observations provide a panoramic overview of the platelet transcriptome, while additionally providing insights into the molecular pathways regulating platelet (and/or megakaryocyte) function in normal and pathological conditions.

## Materials and methods

### Reagents and supplies

*Thermus aquaticus* (Taq) polymerase was purchased from (Roche, Indianapolis, IN). T4 DNA ligase was purchased from Invitrogen (Carlsbad, CA), and restriction enzymes were from New England Biolabs (Beverly, MA), except for *MmeI*, which was obtained from the Center for Technology Transfer (Gdansk, Poland). All oligonucleotides were synthesized on an Applied Biosystems (Foster City, CA) 3-channel synthesizer and are listed in Table 1. Monoclonal antibodies used for flow cytometric analysis included the FITC (fluorescein isothiocyanate)-conjugated anti-CD41 (αIIbβ3) immunoglobulin G1 (IgG1; Immunotech, Miami, FL); phycoerythrin (PE)-conjugated anti-glycophorin (IgG2; Becton Dickinson Pharmingen, San Diego, CA); and peridinin chlorophyll protein (PERCP)-conjugated anti-CD45 (IgG1; Becton Dickinson Pharmingen).

### Platelet isolation, purification, and immunodetection

All human subjects provided informed consent for an IRB (Institutional Review Board)-approved protocol completed in conjunction with the General Clinical Research Center at Stony Brook University Hospital. Peripheral blood (20 mL) from healthy volunteers drawn into 2 mL of 4% sodium citrate (0.4% vol/vol final concentration) was used to isolate erythrocytes by differential centrifugation (1500g) or to isolate pure leukocytes by density gradient centrifugation as previously described.<sup>16</sup> Platelets collected from healthy volunteers by apheresis were used within 24 hours of collection. After addition of 2 mM EDTA (ethylenediaminetetraacetic acid), apheresis-derived platelets from a single donor were centrifuged at 140g for 15 minutes at 25°C. To minimize leukocyte contamination, only the upper 9/10 of the platelet-rich plasma (PRP) was

used for gel filtration over a BioGel A50M column (1000 mL total volume) equilibrated with HBMT (HEPES-buffered modified Tyrodes buffer; 10 mM HEPES (N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid) pH 7.4, 150 mM NaCl, 2.5 mM KCl, 0.3 mM NaH<sub>2</sub>PO<sub>4</sub>, 12 mM NaHCO<sub>3</sub>, 0.2% bovine serum albumen [BSA], 0.1% glucose, 2 mM EDTA). Gel-filtered platelets (GFPs) were subsequently filtered through a 5-μm nonwetting nylon filament filter (BioDesign, Carmel, NY) at 25°C and harvested by centrifugation at 1500g for 10 minutes at 25°C. Platelets were gently and thoroughly resuspended in 10 mL HBMT buffer and incubated with 120 μL murine monoclonal anti-CD45 antibody conjugated to magnetic microbeads (Miltenyi Biotec, Bergisch Gladbach, Germany) on a rotating platform for 45 minutes at 25°C. Magnetic separation columns were used to capture CD45<sup>+</sup> cells (leukocyte fraction) by positive selection (MACS II; Miltenyi Biotec). Purified platelets were concentrated by centrifugation at 1500g and immediately used for total RNA isolation.

The efficiency of platelet purification was documented at each step by flow cytometry.<sup>17</sup> Briefly, aliquots containing 2 × 10<sup>6</sup> platelets were incubated with saturating concentrations of FITC-conjugated anti-CD41, PE-conjugated anti-glycophorin, and PERCP-conjugated anti-CD45 for 15 minutes in the dark at 25°C, washed with phosphate-buffered saline (PBS), and fixed in PBS/1% formalin. Samples were analyzed using a FACScan (fluorescence-activated cell sorter scan) flow cytometer (Becton Dickinson) using CELLQuest software designed to quantify the number of CD45<sup>+</sup> and glycophorin-positive events in the sample (expressed as the number of events per 100 000 CD41<sup>+</sup> events). For some experiments, fixed platelets were permeabilized with 0.1% Triton-X/PBS for 30 minutes at 25°C prior to the addition of primary antibodies, all as previously described.<sup>17</sup>

Platelet protein detection was completed by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) and immunoblot analysis as previously described, using the species-specific horseradish peroxidase-conjugated secondary antibody and enhanced chemiluminescence.<sup>18</sup> Antibodies included the anticlustarin monoclonal antibody (Quidel, Santa Clara, CA; 1:1000 primary and 1:10 000 secondary) and the apineurogranin rabbit polyclonal antibody (Chemicon International, Temecula, CA; 1:1000 primary and 1:10 000 secondary).

### Molecular analyses and microarray profiling

Purified, individual cell fractions were resuspended in 10 mL Trizol reagent (Invitrogen), transferred into diethylpyrocarbonate (DEPC)-treated Corex (Springfield, MA) tubes, and serially purified and precipitated by using isopropanol essentially as previously described.<sup>18</sup> Total cellular RNA was harvested by centrifugation at 12 500g for 20 minutes at 4°C, washed 2 times with 75% ethanol (10 mL/tube), and resuspended in 100 μL DEPC-treated water. Platelet mRNA quantitation was performed by using fluorescence-based real-time PCR (polymerase chain reaction) technology (TaqMan Real-Time PCR; Applied Biosystems, Foster City, CA). Oligonucleotide primer pairs were generated by using Primer3 software ([www.genome.wi.mit.edu](http://www.genome.wi.mit.edu)), designed to generate approximately 200-base pair (bp) PCR products at the same annealing temperature, and are outlined in Table 1. Purified platelet mRNA (4 μg) was used for first-strand cDNA synthesis using oligo(dT) and SuperScript II reverse transcriptase (Invitrogen). For real-time reverse transcription (RT)-PCR analysis, the RT reaction was equally divided among primer pairs and used in a 40-cycle PCR reaction for each target gene by using the following cycle: 94°C for 30 seconds, 55°C for 30 seconds, 72°C for 1 minute, and 71°C for 10 seconds (40 cycles total). mRNA levels were quantified by monitoring real-time fluorometric intensity of SYBR green I. Relative mRNA abundance was determined from triplicate assays performed in parallel for each primer pair and was calculated by using the comparative threshold cycle number (ΔCt method) as previously described.<sup>18</sup>

Gene expression profiles were completed by using the approximately 12 600-probe set HG-U95Av2 gene chip (Affymetrix, Santa Clara, CA). Total cellular RNA (5 μg) was used for cDNA synthesis by using SuperScript Choice system (Life Technologies, Rockville, MD) and an oligo(dT) primer containing the T7 polymerase recognition sequence (Primer S1; Table 1), followed by cDNA purification using GFX spin columns. In vitro transcription was completed in the presence of biotinylated ribonucleotides by using a BioArray HighYield RNA Transcript

Table 1. Oligonucleotide primers

Primer	Gene and primer direction	Sequence (5' - 3')	Nucleotide Position
S1	Oligo (dT)	5'-Bn-GGCCAGTGAATGTAAACCACTCCTATAGGAGGCGG- (dT) <sub>24</sub> -3'	—
Cassette A	SAGE	5'-TTTGATTTCCTGGTCGAGTCACTAGGCTTAATCCGACATG-3' 3'-CCTAAACGACCACTCATGTTGATCCGATAAGGCTP-5'	—
Cassette B	SAGE	5'-pTTCATGCGGAGACGTCGCGCACTAGTGTGCAACTGACTA*-3' 3'-NNAAGTACCCCTCTGCAGGCGGTGATCACGGTTGACTGAT-5'	—
S2	SAGE	5'-Bn-GGATTTCCTGGTCGAGTCA-3'	—
S3	SAGE	5'-Bn-TAGTCAGTGCCGCACTAGTGGC-3'	—
GP4	Glycoprotein IIB [F]	5'-AGGGCTTTGAGAGACTCATCTGTA-3'	2094-2117
GP5	Glycoprotein IIB [R]	5'-ACAATCTTGTCTTTCGATTCTG-3'	2301-2279
GP6	Glycoprotein IIIA [F]	5'-TATAAAGAGGCGACGCTACCTTC-3'	2335-2358
GP7	Glycoprotein IIIA [R]	5'-CACTTCCACTACTGACATTTCTCC-3'	2532-2509
PAR18	PAR1 [F]	5'-AATGTGCTGCTCATATGGAACA-3'	2585-2608
PAR19	PAR1 [R]	5'-CCCAATGTTCAAACTCTTTAGC-3'	2776-2753
SR8	16S rRNA [F]	5'-TGCAAGGATGACATATCACTTGT-3'	2586-2609
SR9	16S rRNA [R]	5'-GTTTAGGACCTGTGGTTTGTAG-3'	2785-2762
NADH10	NADH2 [F]	5'-CTAGCCCCCATCTCAATCATATAC-3'	4875-4898
NADH11	NADH2 [R]	5'-AATGCTTATGTTAGGTTGTACGG-3'	5075-5052
THYM12	Thymosin β4 [F]	5'-AAGACAGAGACGCAAGAGAAAAT-3'	135-158
THYM13	Thymosin β4 [R]	5'-GCAGCAGCTATTAACTTGAT-3'	336-313
CLUS14	Clusterin [F]	5'-CCACAGAAATCATACGAGAGG-3'	1006-1028
CLUS15	Clusterin [R]	5'-CGTTATATTCTGCTCAACCTCT-3'	1222-1199
NRG16	Neurogranin [F]	5'-GCCCTTTTAGTATGTTCTGAGTC-3'	1351-1374
NRG17	Neurogranin [R]	5'-TTTCTTTAAGTGAGTGCTTGG-3'	1567-1544
TCR18	β-chain [F]	5'-CCACACTATGTGTTTGGTATCGT-3'	131-153
TCR19	β-chain [R]	5'-CTAGCACTGCAGATGTAGAGCT-3'	332-310
CD4520	CD45 [F]	5'-GCTCAGATGGACAGTA-3'	3771-3788
CD4521	CD45 [R]	5'-CACACCATACACATACA-3'	4280-4261

[F] indicates forward (sense) strand; [R], reverse (antisense) strand; Bn, biotin; p, a phosphorylated 5' end (cassettes A and B); underlining, *MspI* sites in cassettes A and B; arrows, corresponding sequence for S2 and S3 within cassettes A and B, respectively; bold, the *MspI* site; and N, A, C, T, or G, nucleotide position based on the following accession numbers: glycoprotein IIB (J02764), glycoprotein IIIA (M35999), PAR1 (M62424), 16S rRNA and NADH2 (NC\_001807), thymosin β4 (M17733), clusterin (M25915), neurogranin (X99076), TCR β-chain (AF043182), CD45 (Y00638).

\*Indicates an amino-modified 3' end in both cassettes; —, not applicable.

Labeling Kit (Enzo Diagnostics, Farmingdale, NY), and, after metal-induced fragmentation, 15 μg biotinylated cRNA was hybridized to the JIG-U95Av2 oligonucleotide probe array for 16 hours at 45°C. After washing, the cRNA was detected with streptavidin-phycoerythrin (Molecular Probes, Eugene, OR) and analysis was completed by using a Hewlett-Packard Gene Array Scanner (Affymetrix). The average difference value (AD) for each probe set was quantified using MAS 4.0 software (Affymetrix), calculated as an average of fluorescence differences for perfectly matched versus single-nucleotide mismatched 25-mer oligonucleotides (16 to 20 oligonucleotide pairs per probe set). The software is designed to exclude "positive calls" in the presence of high average differences with associated high mismatch intensities.

#### SAGE profiles

Platelet SAGE libraries were generated essentially as previously described,<sup>12</sup> modified as outlined in Figure 1 for the use of *MspI* as the tagging enzyme.<sup>13</sup> This type IIS restriction enzyme cleaves 20 of 18 bp past its nonpalindromic (TCCRAC) recognition sequence, thereby generating longer tags (21- or 22-mer) than those obtained using *BsmI* as the standard tagging enzyme (13-14 bp tags). These longer *MspI*-generated tags potentially provide for more definitive "tag-to-gene" identification and are particularly useful in characterizing expression patterns in the absence of complete genomic sequence data (comprehensive methods detailed in Dunn et al<sup>14</sup>). Briefly, poly(A) mRNA was isolated from 10 μg total platelet RNA using the oligo-d(T) S1 primer conjugated to magnetic beads (Dyna Beads, Biotec, Lake Success, NY), followed by cDNA synthesis using SuperScript II reverse transcriptase (Invitrogen). The cDNA was then digested

with the restriction enzyme *MspI* (anchoring enzyme), ligated to cassette A using T4 DNA ligase, and, after the beads were extensively washed, the cDNA was digested with *MspI* to release the tags from the beads. After purification, tags were ligated to degenerate cassette B linkers (specifically

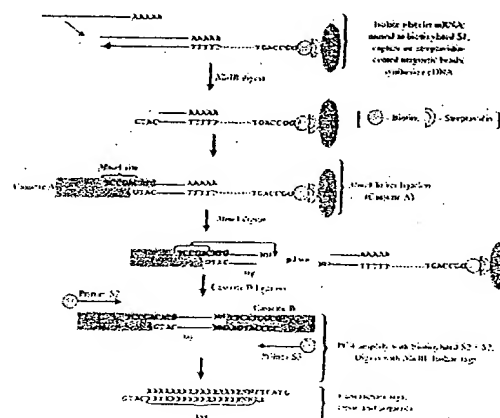


Figure 1. Schema outlining the modified SAGE protocol used in platelet analyses. The final tags are flanked by the *MspI* (anchoring enzyme) CATG sequence, thereby providing tag-to-gene identification when exported to a relational database (refer to "Bioinformatic analyses" and Table 1 for details).

designed to anneal to the nonuniform *MmeI* overhangs), and PCR-amplified using biotinylated primers S2 and S3 for 30 cycles (95°C for 30 seconds; 58°C for 30 seconds; 72°C for 30 seconds) using Platinum Taq DNA polymerase (Gibco-BRL). A fraction (20%) of the pooled PCR products were then subjected to one round of linear amplification using primer pair S2/S3, followed by a second round of 25 amplifications using primer S2 alone (95°C for 30 seconds; 58°C for 30 seconds; 72°C for 30 seconds). Primer S3 was subsequently added for one cycle (95°C for 2.5 minutes; 58°C for 30 seconds; 72°C for 5 minutes); the latter steps were collectively adapted to exclude heteroduplex formation.<sup>18</sup> Unincorporated primers were removed by incubation with 200 U *Escherichia coli* exonuclease I for 60 minutes at 37°C. PCR products were then pooled and digested with *MolII* to release tags, and biotinylated linker arms were cleared using streptavidin-coated immunoadfinity magnetic beads (Dynal Biotech). Tags were concatamerized using 5 U/μL T4 DNA ligase, and products more than 100 bp were isolated by size-fractionation in low-melting agarose gels. The DNA was purified by GFX spin columns, and the concatamers were cloned into the *SphI* site of pZero (Invitrogen). After transformation into *E. coli* TOP10 cells, recombinant clones were isolated and sequenced in 96-well microtiter plates using an ABI 377 sequencer and ABI Prism BigDye terminator chemistry (Perkin-Elmer Applied Biosystems, Branchburg, NJ).

#### Bioinformatic analyses

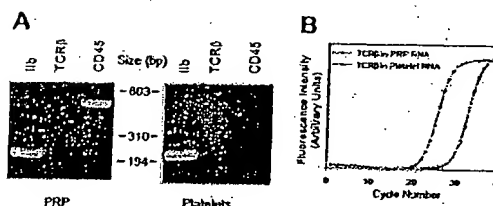
Functional grouping of genes determined to be present by Affymetrix MAS 4.01 software was performed using a dChip program linked to the National Center for Biotechnology LocusLink, which is an annotated reference database for genes and their postulated functions.<sup>20</sup> Of the approximately 12 600 probe sets represented on the Affymetrix HG-U95Av2 Gene chip, functional annotations exist for approximately 8100 with the remainder categorized as unknown. Microarray data were visualized and analyzed using BRB-ArrayTools software (Version 2.1), kindly developed and provided by Dr Richard Simon and Amy Peng ([lims.nci.nih.gov/BRB-ArrayTools.html](http://lims.nci.nih.gov/BRB-ArrayTools.html)). A logarithmic (base 2) transformation was applied to the average difference values for individual data sets for determination of microarray concordancies. Discordancy was defined as a 2-log difference in the maximum log intensities between individual experiments.

SAGE tags were extracted by using in-house SAGE software uniquely modified to identify *MmeI* tags. The software ensures that only unambiguous 21- to 22-bp tag sequences are extracted for transcript profiling. Tags with ambiguities (Ns), lengths other than 21 or 22 bp, or with ambiguous orientations were extracted to separate files for manual editing or further examination. Finalized data were exported to a relational database for tag quantification and genetic identification.<sup>20</sup>

## Results

### Platelet purification

To ensure that the RNA profiles accurately represented those of circulating blood platelets, a number of complementary methods were implemented to remove contaminating nucleated leukocytes. Purification methods incorporating gel filtration, a 5-μm leukocyte reduction filter, and magnetic CD45 immunodepletion allowed for the cumulative enrichment of highly purified platelets. The efficacy of this purification method was initially established by using peripheral blood platelet-rich plasma as the starting material. The final product contained no more than 3 to 5 leukocytes per  $1 \times 10^5$  platelets as determined by parallel flow cytometric analysis, representing an approximate 450-fold reduction of nucleated leukocytes. These results correlated well with molecular evidence for leukocyte depletion as determined by RT-PCR using both CD45 and T-cell receptor  $\beta$ -chain (TCR $\beta$ ) primers (see Figure 2). Because the total RNA yield from peripheral blood platelets was insufficient for microarray studies, we adapted the protocol to platelet apheresis donors with nearly identical final purity (Figure



**Figure 2.** Determination of platelet purity. (A) Total cellular RNA (1.8 μg) from platelet-rich plasma (PRP) or purified platelets from a single apheresis donor were analyzed by RT-PCR (35 cycles) using oligonucleotide primers specific for glycoprotein IIb (GP1Ib), T-cell receptor  $\beta$ -chain (TCR $\beta$ ), or CD45; 10 μL of the 50 μL reactions were analyzed by ethidium-stained agarose gel electrophoresis. Minimal to no TCR $\beta$  gene product was visually evident only in PRP. Size markers corresponding to *HaeIII*-restricted  $\lambda$ X174 DNA are shown. (B) Real-time RT-PCR was completed by using 1.8 μg total RNA and TCR $\beta$ -specific oligonucleotide primers optimized for quantitative analysis by real-time PCR.<sup>18</sup> On the basis of parallel determinations using RNA isolated from known amounts of purified leukocyte standards, the leukocyte-depletion protocol represents an approximate 2.5-log purification from the starting PRP. Results are representative of one complete set of experiments repeated on 2 separate occasions, and data points represent the mean from triplicate wells, with standard errors of the mean (SEM) less than 1% (not shown).

2). The platelet recovery was nearly 65% of the starting material, yielding approximately  $2.3 \times 10^{11}$  platelets from an initial apheresis pack containing approximately  $3.6 \times 10^{11}$  platelets. The bulk of the losses occurred during the initial centrifugation and filtration steps. The purification protocol was less effective at removing erythrocytes, although there were less than 50 glycophorin-positive cells per  $1 \times 10^5$  platelets after the final purification step. Nonetheless, these cells represent unlikely sources for contaminating cellular RNA (see "Cellular microarray analysis" below).

### Cellular microarray analysis

The purified platelet RNA was sufficient for microarray studies and was used for cRNA generation and hybridization to the Affymetrix HG-U95Av2 GeneChip. The anatomic profile of platelet RNAs from 3 healthy male donors was determined by using Affymetrix software. Of the 12 599 probe sets imprinted onto the chip, a maximum of 2147 (17%) transcripts were computationally identified as "present" by the Affymetrix software, 152 (1.2%) were equivocal, and nearly 82% were absent. As a fraction of the total genes present on the chip, the percentage of platelet-expressed genes (15%-17%) was generally lower than that obtained from other human cell types in which 30% to 50% of genes are present as determined by Affymetrix software (J. Schwedes, personal communication, May 2002). The "limited number" of platelet-expressed transcripts presumably reflects the lack of ongoing gene transcription in the anucleate platelet. Because less than 1% of circulating red blood cells contain residual RNA, it is unlikely that any of these transcripts are erythrocyte derived, although this was formally addressed by isolating total cellular RNA from 20 mL of whole blood (corresponding to an ~3-log fold excess of erythrocytes than that identified in our final sample). The total cellular yield of RNA from this starting material was approximately 250 ng, suggesting that less than 1 ng erythrocyte-derived RNA was present in the purified platelet preparations. Despite this, however, both  $\alpha$ - and  $\beta$ -globin transcripts—along with both the ferritin heavy and light chains—were identified as abundant transcripts (Table 2). Although the most parsimonious explanation would be residual contaminating reticulocytes, this is not supported by our erythrocyte contamination estimates, and their significance remains unresolved.

As a means of better dissecting the molecular anatomy of the platelet, expressed genes were grouped on the basis of assigned

Table 2. Top 50 human platelet-expressed genes

Accession no.	Gene symbol	AD values, range*	Gene transcript†	Leukocyte expression‡
M17733	TMSB4X	140 142-307 852	Thymosin $\beta$ 4 mRNA, complete cds	+
X99076	NRGN	101 510-148 279	Neurogranin gene	+
M25079	HBB	40 839-229 556	$\beta$ -globin mRNA, complete cds	+
M25915	CLU	84 720-140 246	Complement cytolysis inhibitor (clusterin) complete cds	-
J04755	FTHP1	82 980-148 621	Femlin H processed pseudogene, complete cds	-
D78361	OAZ1	73 098-118 140	mRNA for ornithine decarboxylase antizyme	-
X04409	GNAS	77 761-94 781	mRNA for coupling protein G(s) $\alpha$ -subunit (alpha-S1)	-
M25897	PF4	62 811-126 908	Platelet factor 4 mRNA, complete cds	-
AB021288	B2M	61 689-108 921	$\beta$ 2-microglobulin	+
X00351	ACTB	25 143-73 775	mRNA for $\beta$ -actin	-
D21261	TAGLN2	76 687-101 931	mRNA for KIAA0120 gene	+
AL031670	FTLL1	69 865-99 966	Ferritin, light polypeptide 1	+
U59632	GPIIB	41 404-110 328	Platelet glycoprotein IIb chain mRNA	-
M21121	CCL5	47 308-106 399	T-cell-specific protein (RANTES) mRNA, complete cds	-
X13710	GPX1	41 318-96 878	Unspliced mRNA for glutathione peroxidase	-
J00153	HBA1	21 326-144 201	Alpha globin gene cluster on chromosome 16	+
M22919	MYL6	46 337-106 833	Nonmuscle/smooth muscle alkali myosin light chain gene	+
L20941	FTH1	52 787-74 763	Ferritin heavy chain mRNA, complete cds	-
J03040	SPARC	51 156-74 261	SPARC/osteonectin mRNA, complete cds	-
X56009	GNAS	45 543-72 096	GSA mRNA for $\alpha$ subunit of GsGTP binding protein	-
X58536	HLA	31 183-82 613	mRNA for major HLA class I locus C heavy chain	+
M54995	PPBP	46 571-67 169	Connective tissue activation peptide III mRNA	-
U34995	GAPD	35 095-70 250	Normal keratinocyte subtraction library mRNA, clone H22a	+
L40399	MLM3	32 107-73 364	Clone zap112 (mutL protein homolog 3) mRNA	-
X77548	NCQA4	31 452-61 036	cDNA for RFG (RET proto-oncogene RET/PTC3)	-
U90551	H2AFL	35 086-51 892	Histone 2A-like protein (H2AFL) mRNA	-
M11353	H3F3A	31 614-55 813	H3.3 histone class C mRNA	-
Z12962	RPL41	36 003-54 853	mRNA for homologue to yeast ribosomal protein L41	+
X06956	TUBA1	20 988-61 798	HALPHA 44 gene for $\alpha$ -tubulin	-
AB028950	TLN1	24 571-58 611	mRNA for KIAA 1027 protein	-
Y12711	PGRMC1	33 680-43 174	mRNA for putative progesterone binding protein	-
M16279	MIC2	30 894-48 166	Integrated membrane protein (MIC2) mRNA	-
D78577	YWHAH	24 785-50 437	Brain 14-3-3 protein $\beta$ -chain	-
AF070585	TOP3B	20 027-67 945	Clone 24675, unknown cDNA	+
AA524802	Unknown	23 846-39 481	CDNA, IMAGE clone 954213	-
AB009010	UBC	28 745-38 389	mRNA for polyubiquitin UbC	+
X57985	H2AFO	21 678-52 108	Genes for histones H2B.1 and H2A	-
X54304	MLCB	25 733-34 109	mRNA for myosin regulatory light chain	-
M14539	F13A1	23 691-48 474	Factor XIII subunit $\alpha$ -polypeptide mRNA, 3' end	-
AJ540958	Unknown	24 872-41 118	cDNA, PEC 1.2_15_HOI, 5' end / clone	-
AL050396	FLNA	13 634-55 235	cDNA DKFZp586K1720	-
X56841	HLA-E	12 890-49 327	Nonclassical MHC class I antigen gene	-
M26252	PKM2	15 450-47 786	TCB (cytosolic thyroid hormone-binding protein)	-
M14630	PTMA	19 314-45 088	Prothymosin alpha mRNA	-
AF045279	RGS10	19 156-34 243	Regulator of G protein signaling 10 mRNA	-
AA477898	Unknown	16 863-44 756	cDNA, Z834108.115' end	-
X95404	FLI1	15 216-37 456	mRNA for nonmuscle type cofilin	-
M34480	ITGA2B	8 627-45 495	Platelet glycoprotein IIb (GPIIb) mRNA	-
Z83738	H2BFE	18 001-31 306	H2BFE gene	-
L19779	H2AFO	17 319-38 951	Histone H2A.2 mRNA, complete cds	-

\*Gene expression quantifications were calculated as the average difference (AD) value (matched versus mismatched oligonucleotides) for each probe set using Affymetrix GeneChip software, version 4.01. The range of values from 3 distinct platelet microarrays is shown; the normalization value for all microarray analyses was 250.

†Transcripts are rank-ordered (highest to lowest) using BRB-ArrayTools software by log-intensities of AD values obtained from 3 different healthy donors; 33 of the top 40 transcripts were listed among the top 50 in all 3 microarray sets.

‡Leukocyte expression was determined by microarray analysis using purified peripheral blood leukocytes, followed by construction of rank-intensity plots for comparison to platelet top 50 transcripts.<sup>20</sup> Top leukocyte-derived transcripts identified within the ranked top 50 platelet transcripts are depicted by a (+) present, or (-) absent, cds indicates coding sequence.

gene annotations, and this analysis was used to provide a panoramic definition of the platelet transcriptome. Of the genes that could be cataloged within assigned "clusters," those involved in metabolism (11%) and receptor/signaling (11%) represented the largest groups. Also evident in these analyses is the relatively large percentage of genes involved in functions unrelated to these key groups (ie, miscellaneous, 25%), and the overrepresentation of genes with unknown function (32%) as annotated by Affymetrix

and RefSeq databases.<sup>21</sup> These results identify a vast array (nearly one half) of platelet genes (and gene products) that presumably have important, but poorly characterized functions, in platelet and/or megakaryocyte biology.

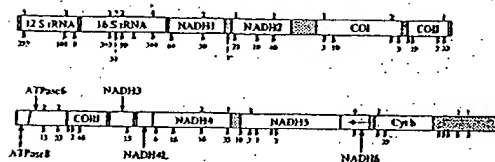
Although microarray analysis is not truly quantitative, rank-ordering using the mean log-intensities from 3 independent microarray analyses allowed for the categorization of the top platelet transcripts (Table 2). Computational analyses demonstrated that



only 10 of the top 100 genes were discordant among the 3 platelet microarrays, although 71 of 100 genes were discordant between platelet and leukocyte arrays. An inventory of the top 50 platelet genes is listed in Table 2, which also delineates those found to be highly expressed in peripheral blood leukocytes by parallel microarray experiments with this purified cellular fraction (data not shown). Further analysis of these cell subsets demonstrated that approximately 25% ( $n = 547$ ) of the total platelet transcripts were platelet restricted. Furthermore, only 10 of the 50 most highly expressed genes were found to overlap, confirming the distinct cellular profiles of each transcriptome. Of the 12 overlap genes, 3 corresponded to globin or ferritin chains (again suggesting the presence of contaminating reticulocytes in both purified fractions), and another 4 were involved in actin cytoskeletal reorganization and human leukocyte antigen (HLA) expression, gene products that regulate critical functions in both cell types. Given the importance of cytoskeletal reorganization in downstream platelet activation events, it is not unexpected that components of the actin machinery system would demonstrate prominent transcript expression. Previous estimates suggest that 20% to 30% of the total platelet proteome is comprised of actin with other components such as actin-binding protein, myosin, and talin accounting for an additional 2% to 5% of the total protein.<sup>1,22</sup> The mRNAs encoding the actin-related machinery are overrepresented in our microarray analysis, with 8 such transcripts found among the 50 highest platelet-expressed genes. Interestingly thymosin  $\beta_4$  demonstrated the highest expression pattern. In unstimulated platelets, 30% to 40% of actin is polymerized as F-actin,<sup>22</sup> whereas the balance of actin monomers (G-actin) are polymerization inhibited by sequestering proteins such as profilin (100  $\mu$ M) and thymosin  $\beta_4$  (600  $\mu$ M).<sup>23</sup> The high thymosin  $\beta_4$  transcript expression not only correlates with its known abundance in platelets but also supports the importance of actin inhibitory proteins in maintaining the nonstimulated state of circulating platelets.

#### Platelet SAGE analyses

Although these initial studies identified the distribution and relative expression patterns of the genes within the Affymetrix data set, they do not allow for analyses of genes that are unrepresented by these oligonucleotide chips. Unlike closed microarray profiling strategies, SAGE is an open architectural system that is ideally suited for novel gene and pathway identification. Accordingly, the platelet RNA used for microarray studies was used for platelet SAGE. A total of 2033 tags were initially cataloged, of which 1800 (89%) corresponded to mitochondrial-derived genes. These results were quite different from those obtained by microarray analyses, but the discrepancy can be resolved by the nonrepresentation of the mitochondrial genome on the gene chip. The mitochondrial genome is a compact approximately 16.6-kilobase (kb) sequence encoding 13 genes and 2 ribosomal subunits.<sup>24</sup> Primary mitochondrial transcripts are polycistronic and typically contain premature termination or unpredictable splice sites, resulting in multiple polyadenylated transcripts from individual genes.<sup>24,25</sup> Indeed, the overall distribution of platelet-derived mitochondrial SAGE tags is quite similar to that found in muscle.<sup>22</sup> All 13 genes containing *Nla*III sites were detected, whereas neither of the non-*Nla*III-containing genes were identified (nicotinamide adenine dinucleotide [NADH] dehydrogenase subunit 4L and adenosine triphosphatase [ATPase] 8). Most of the tags were from the 16S and 12S ribosomal RNAs—which collectively accounted for 68% of the total mitochondrial tags—with the fewest tags represented by NADH dehydrogenase subunits 3, 5, 6, and cytochrome c oxidase I



**Figure 3.** Schema of the mitochondrial genome with SAGE tag distributions (only tags with identical matches are displayed). The abundance of the SAGE tags ( $n = 1800$ ) at individual *Nla*III sites (arrows) within the mitochondrial heavy strand is shown on the bottom, whereas those tags corresponding to the mitochondrial light strand are delineated above the arrows (the presence of an unaccompanied arrow implies no SAGE tags at that *Nla*III site). The gene products of mt-DNA (RefSeq accession no. NC\_001807) are delineated by the open rectangles, whereas stippled boxes represent tRNA genes and control regions (the single tag represented by the [•] refers to mitochondrial transfer RNA-serine). Note that NADH6 is encoded by the light strand and that there are no *Nla*III sites within the ATPase8 gene segment. COX1, cytochrome c oxidase subunit; Cyt b, cytochrome b.

(Figure 3). The NADH dehydrogenase subunit 6 RNA is the only mRNA encoded by the light (L) strand of mitochondrial DNA and was the least abundantly detected transcript.

The unusually high preponderance of mitochondrial-derived genes is not inconsistent with the known enrichment of these genomes in human platelets,<sup>1,24</sup> and presumably reflects persistent transcription from the mitochondrial (mt) genome in the absence of nuclear-derived transcripts. This overrepresentation of mtDNA in platelets is considerably greater than that of its closest cell type (skeletal muscle), in which mt genomes represent approximately 20% to 25% of all SAGE tags.<sup>25</sup> Interestingly, the energy metabolism of platelets is not dissimilar from that of skeletal muscle, both cell types actively using glycolysis and large amounts of glycogen for ATP generation.<sup>26</sup> Like muscle, platelets are metabolically adapted to rapidly expend large amounts of energy required for aggregation, granule release, and clot retraction. Similar to the situation in all eukaryotic cells, platelet mitochondria represent the primary source of ATP, which is generated from oxidative phosphorylation reactions occurring within these organelles. Mitochondria are also responsible for most of the toxic reactive oxygen species generated as by-products of oxidative phosphorylation and are central regulators of the apoptotic process in other cellular types. The mtDNA encodes polypeptides found within 4 of the 5 multifunctional complexes that regulate oxidative phosphorylation within the platelet mitochondria.<sup>27</sup> Whether the continued generation of these polypeptides has a role in platelet energy metabolism and/or the apoptotic mechanisms regulating platelet survival remains speculative, although not inconsistent with our observations.

#### Comparative analysis of SAGE and microarray transcript abundance

Complete SAGE libraries require the sequencing of up to 30 000 tags for an exhaustive cataloging of individual mRNAs, especially those with limited copy numbers.<sup>13,28</sup> Given the preponderance of mt-derived transcripts, comparable sampling would have required sequence analysis of nearly 300 000 SAGE tags, an inordinate number for comprehensive analysis of the platelet transcriptome. For platelets, alternative methodologies incorporating subtractive SAGE will be required for more comprehensive transcript profiling.<sup>29</sup> Our initial sampling of nonmitochondrial genes remains informative, however, and entirely consistent with the results of platelet microarray studies. As shown in Table 3, SAGE tags for the genes encoding thymosin  $\beta_4$ ,  $\beta_2$ -microglobulin, neurogranin, and the platelet glycoprotein Ib polypeptide were among the most frequently identified platelet genes, similar to the rank-ordered results determined by microarray analysis. To formally confirm the

Table 3. SAGE-identified nonmitochondrial tags

Frequency	CATG + SAGE tags*	Accession no.†	Gene	Microarray‡
26	GTGTGGTTAATCTGGT	NM_004048.1	$\beta$ 2-microglobulin (B2M), mRNA	PPP
21	TTGGTGAAGGAAGAAGT	NM_021109.1	Thymosin $\beta$ 4; X chromosome (TMSB4X), mRNA	P
8	AGCTCCGACGCCAGGTC	NM_002620.1	Platelet factor 4 variant 1 (PF4V1), mRNA	P
8	AGCTCCGACGCCGGGTT	NM_002619.1	Platelet factor 4 (PF4), mRNA	P
7	TGTATAAAGACAACCTC	NM_002704.1	Proplatelet basic protein ( $\beta$ -thromboglobulin)	Pp
5	GGGCACAATGCGGTCCA	NM_000407.1	Glycoprotein Ib $\beta$ polypeptide, mRNA	P
3	AGGTAATAAAGGTAAT	NM_003512.1	H2A histone family, member L (H2AFL), mRNA	P
3	AGTGGCAAGTAAATGGC	NM_021914.2	Cofilin 2 (muscle) (CFL2), mRNA	N/A
3	TGACTGTGCTGGGTTGG	NM_006176.1	Neurogranin (protein kinase C substrate, RC3) mRNA	P
3	TTGGGGTTTCCTTTACC	NM_002032.1	Ferritin, heavy polypeptide 1 (FTH1), mRNA	P
2	CCCTTGTGACTACCTAT	NM_025158.1	Hypothetical protein FLJ22251 (FLJ22251), mRNA	N/A
2	CCTGTAAACCCAGCTAC	NM_032779.1	Hypothetical protein FLJ14397 (FLJ14397), mRNA	N/A
2	CTTGAGTCCCAGCTAC	NM_017962.1	Hypothetical protein FLJ20825 (FLJ20825), mRNA	N/A

\*Unique tags identified more than once.

†Refers to the RefSeq accession no.‡Note that this number does not necessarily correspond to the accession no. provided by Affymetrix software annotations (Table 1).

‡Presence (P) or absence (A) is based on results from 3 distinct platelet microarray experiments. Capitalized "P" designates a gene that is in the top 50 on all 3 microarray experiments, whereas small "p" designates those transcripts not in the top 50. Two of the genes ( $\beta$ 2-microglobulin and  $\beta$ -thromboglobulin) are represented by 3 and 2 probe sets, respectively, on the HG-U95Av2 gene chip; for  $\beta$ 2-M, all 3 probe sets were in the top 50 genes, whereas for thymosin  $\beta$ 4 1 of 2 was in the top 50 for all experiments (the other probe set was in the top 75 for all experiments). N/A indicates oligonucleotide not present on Affymetrix HG-U95Av2 gene chip.

results independently obtained by SAGE and microarray analysis, quantitative RT-PCR was completed by using oligonucleotide primers specific for 2 abundant mitochondrial transcripts, 16S rRNA and NADH2 thymosin  $\beta$ 4 (high-abundance by microarray and SAGE), 2 incompletely characterized high-abundance transcripts (neurogranin and clusterin; see "Protein immunoanalysis of platelet clusterin and neurogranin"), a low-abundance transcript (T-cell receptor  $\beta$ -polypeptide), and the genes encoding proteins with well-established quantitative determinations (ie, glycoprotein  $\alpha_{IIb}\beta_3$  [ $\sim$ 50 000 receptors/platelet]; protease-activated receptor-1 (PAR1) [ $\sim$ 1800 receptors/platelet]).<sup>1</sup> As shown in Figure 4, these analyses reveal excellent concordance between SAGE and microarray studies, demonstrating the predominant frequency of the mitochondrial-derived 16S rRNA/NADH2 transcripts, with incrementally lower expression of other transcripts as initially demonstrated by microarray (16S > NADH2 > thymosin  $\beta$ 4 > neurogranin > clusterin >  $\alpha_{IIb}\beta_3$  > PAR1 > TCR $\beta$ ).

Given the small number of nonmitochondrial SAGE tags available for analysis ( $n = 233$ ), limited conclusions can be drawn using traditional (nonsubtraction) platelet SAGE libraries as pre-

sented here. Overall, a total of 126 unique tags were identified, the majority of which (94) were represented only once. Of the total unique tags, nearly one half represented novel genes not present on the Affymetrix U95Av2 GeneChip. Of the genes with unique tags identified more than once, there was excellent concordance with microarray expression analysis, with nearly all of the SAGE tags in Table 3 corresponding to platelet top 75 microarray transcripts. The platelet factor (PF) 4 variant represents a single aberration because this was rank-ordered approximately 350 by microarray, although its SAGE tag frequency was identical to that of the predominant PF4 transcript. The lack of extensive nonmitochondrial SAGE sampling precludes any further extrapolations from this apparent aberration. Of note, a subset of these tags had long poly(A) tracts; although they all corresponded to genes identified in the RefSeq database.<sup>21</sup> We cannot exclude the possibility of a SAGE artifact for this small subset of tags ( $\sim$ 2%, representing 46 of 2033 tags), although the authenticity of the vast majority of tags ( $\sim$ 98%) clearly validates the methodology. These tags are most likely explained by the unique biology of the platelet (ie, mRNA decay in the absence of de novo transcription) or to mRNA degradation occurring during the extensive purification methods. In summary, even with a remarkably limited sampling, the power of this approach in gene identification of relatively abundant and less abundant transcripts is evident. It is clear, however, given the unique molecular anatomy of the platelet (ie, abundance of mitochondrial transcripts), that SAGE adaptations will be required for more comprehensive genetic profiling.<sup>29</sup>

#### Protein immunoanalysis of platelet clusterin and neurogranin

Although most of the "most abundant" transcripts would conform to a priori predictions for platelet-expressed mRNAs, a number of transcripts were identified that had been poorly characterized in human platelets. To further establish the authenticity of highly expressed transcripts such as clusterin and neurogranin, confirmatory protein analyses were completed. As shown in Figure 5, both proteins were clearly detected in purified platelet lysates; furthermore, their cellular platelet distributions conformed to those predicted based on previously proposed functions. Note for example that clusterin—functionally characterized as a complement lysis inhibitor able to block the terminal complement reaction—is primarily expressed on the extracellular platelet membrane.<sup>30</sup>

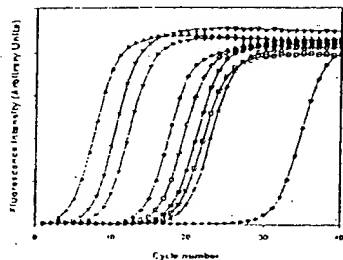
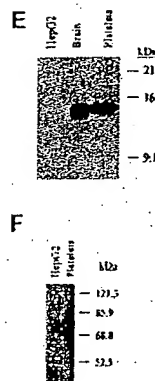
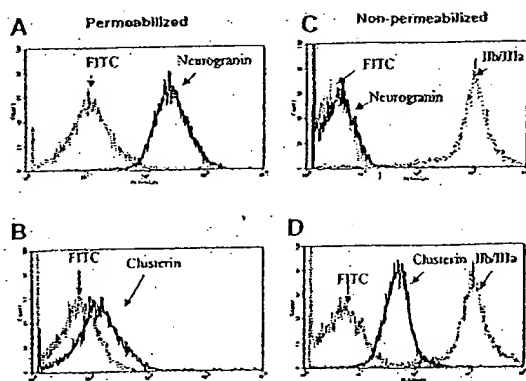


Figure 4. Quantitative real-time RT-PCR analysis of platelet transcripts. Real-time RT-PCR was completed by using purified platelet RNA and oligonucleotide primer pairs specifically designed using Primer3 software to generate similarly-sized ( $\sim$ 200 bp) PCR products, optimized to the same annealing temperature. In graph, (□) represents 16S, (■) represents NADH2, (△) represents thymosin  $\beta$ 4, (▲) represents 16S rRNA, (▽) represents NADH2, (▼) represents thymosin  $\beta$ 4, (○) represents clusterin, (●) represents neurogranin, and (◆) represents TCR $\beta$ . Curves are representative of one complete set of experiments (repeated twice), and line plots reflect average determinations from 3 wells performed in parallel with SEM less than 1% for all data points.



**Figure 5.** Immunoblot analysis of platelet neurogranin and clusterin. (A-D) Gel-filtered platelets were either fixed in 3.7% formaldehyde (nonpermeabilized) or fixed with permeabilization in the presence of 0.1% Triton-X, followed by flow cytometric analysis using anti-clusterin, anti-IIb/IIIa, or anti-neurogranin antibodies and the FITC-conjugated species-specific secondary antibody (in C, the FITC-conjugated anti-rabbit and anti-mouse controls are essentially superimposed). (E-F) Ten micrograms of solubilized HepG2 cells (hepatocyte cell line), human brain, or purified platelet lysates were analyzed by SDS-PAGE,<sup>27</sup> and immunoblot analysis was completed by using 1:1000 dilutions of either anti-neurogranin (18% SDS-PAGE) or anti-clusterin (8% SDS-PAGE) antibodies. The anti-clusterin antibody recognized 2 platelet immunoreactive species under shorter exposure. Although the relative neurogranin and clusterin protein abundances are suboptimally quantified by these analyses, platelet clusterin appears to demonstrate considerable expression when compared with that previously identified in hepatocytes.<sup>21</sup>

Given the importance of complement activation in platelet destruction, the prominent expression of cell-surface clusterin might suggest a role for this protein in normal and pathologic events regulating platelet survival. Interestingly, a clusterin-deficient knockout mouse has been generated that demonstrates enhanced cardiac dysfunction in a model of autoimmune myocarditis.<sup>31</sup> Although these mice apparently have normal baseline hemograms (B. Aronow, personal communication, October 2002), it remains unestablished if they would be predisposed to immune-type thrombocytopenia in systemic models of autoimmunity.

Similarly, the gene encoding an intracellular effector protein that may have key roles in downstream platelet activation events has now been demonstrated to have abundant transcript expression in human platelets. Neurogranin is a highly expressed platelet transcript with its gene product demonstrating a primarily intracellular pattern of distribution. Neurogranin is generally described as a brain-specific,  $Ca^{2+}$ -sensitive calmodulin-binding phosphoprotein that is preferentially expressed in neuronal cell bodies and dendrites.<sup>32,33</sup> It is a specific protein kinase C (PKC) substrate that can also be modified by nitric oxide and other oxidants to form intramolecular disulfide bonds. Both its phosphorylation and oxidation state attenuate its binding affinity for calmodulin.<sup>33</sup> In stimulated platelets, PKC generation is linked to various activation pathways such as calcium-regulated kinases, mitogen-activated protein (MAP) kinases, and receptor tyrosine kinases.<sup>1</sup> Thus, these observations suggest that platelet neurogranin may function as a previously unidentified component of a PKC-dependent activation pathway coupled to one (or more) of these effector proteins.

## Discussion

These data provide documentation for a unique platelet mRNA profile that may provide a tool for analyzing platelet molecular networks. Nonetheless, the molecular analysis of the platelet transcriptome may be confounded by the constant decay of mRNAs in the absence of new gene transcription, a situation that may, for example, limit the identification of low-abundance transcripts. Similarly, because the circulating platelet pool contains

a mixed population of variably aged platelets, a "static" mRNA profile represents an average of this heterogeneous blood pool. Despite these potential limitations, the combination of genomic and proteomic technologies are likely to provide powerful tools for the global analysis of platelet function. Current strategies for cataloging "whole cellular proteomes" are generally accomplished by using 2 developing methodologies: (1) high resolution 2-dimensional polyacrylamide gel electrophoresis (2-DE) with mass spectrometric sequence identification,<sup>34</sup> and (2) microcapillary liquid chromatography with tandem mass spectrometry ( $\mu$ LC-MC/MS).<sup>35</sup> Further modifications of both procedures have been devised for direct comparative studies between 2 cellular proteomes. The introduction of 2-DE differential gel electrophoresis has now made it possible to detect and quantify differences between experimental sample pairs resolved on the same 2-dimensional gel.<sup>36</sup> Likewise, the application of isotope-coded affinity tags to  $\mu$ LC-MC/MS represent a novel means of quantitative analyses between cellular proteomes.<sup>37</sup> The success of both approaches relies on the availability of comprehensive genomic databases and mathematical algorithms for optimal protein identification. Indeed, mathematical modeling studies have demonstrated the need to delineate both protein and mRNA expression levels for optimal definition of intracellular networks.<sup>38</sup> Our data present an initial framework for delineating platelet function by defining the molecular anatomy of human platelets, information that is likely to provide important clues into the dynamic protein interactions regulating normal and pathologic platelet functions. Furthermore, because the platelet transcriptome mirrors the mRNAs derived from precursor megakaryocytes, these analyses may provide insights into the biochemical and molecular events regulating megakaryocytopoiesis and/or proplatelet formation.

## Acknowledgments

We thank Dr Maureen Krause, Jean Wainer, and Lesley Scudder for assistance with some of the experiments; John Schwedes (University DNA microarray facility) with the microarray analysis; and Ms Shirley Murray for manuscript preparation.

## References

- Steinberg P, Hill R. Platelets and megakaryocytes. In: Le R, et al, eds. *Wintrobe's Clinical Hematology*. Baltimore, MD: Williams & Wilkins; 1999.
- Newman P, Gorski J, White G, Gidwitz S, Cretney C, Foster R. Enzymatic amplification of platelet-specific messenger RNA using the polymerase chain reaction. *J Clin Invest*. 1988;82:739-743.
- Kieffer R, Guichard J, Forcet J, Vainchenker W, Breton-Gorius J. Biosynthesis of major platelet proteins in human blood platelets. *Eur J Biochem*. 1987;164:189-195.
- Weyrich A, Dixon D, Pober R, et al. Signal-dependent translation of a regulatory protein, Bcl-2, in activated human platelets. *Proc Natl Acad Sci U S A*. 1998;95:5556-5561.
- Benecke BJ, Ben Zeev A, Perman S. The control of mRNA production, translation and turnover in

- suspended and reattached anchorage-dependent fibroblasts. *Cell*. 1978;14:931-939.
6. Pabla R, Weyrich AS, Dixon DA, et al. Integrin-dependent control of translation: engagement of integrin  $\alpha$ IIb $\beta$ 3 regulates synthesis of proteins in activated human platelets. *J Cell Biol*. 1999;144:175-184.
  7. Chicurel ME, Singer RH, Meyer CJ, Ingber DE. Integrin binding and mechanical tension induce movement of mRNA and ribosomes to focal adhesions. *Nature*. 1998;392:730-733.
  8. Lindemann S, Tolley N, Eyre J, Kraiss L, Mahoney T, Weyrich A. Integrins regulate the intracellular distribution of eukaryotic initiation factor 4E in platelets. *J Biol Chem*. 2001;276:33947-33951.
  9. Brown EJ, Schreiber SL. A signaling pathway to translational control. *Cell*. 1996;85:517-520.
  10. Rinder H, Schuster J, Rinder C, Wang C, Schwesidler H, Smith B. Correlation of thrombosis with increased platelet turnover in thrombocytosis. *Blood*. 1998;91:1288-1294.
  11. Richards E, Baglin T. Quantitation of reticulated platelets: methodology and clinical application. *Br J Haematol*. 1995;91:445-451.
  12. Robinson M, Mackie I, Khair K, et al. Flow cytometric analysis of reticulated platelets: evidence for a large proportion of non-specific labelling of dense granules by fluorescent dyes. *Br J Haematol*. 1998;100:351-357.
  13. Velculescu V, Zhang L, Vogelstein B, Kinzler K. Serial analysis of gene expression. *Science*. 1995;270:484-487.
  14. Zhang L, Zhou W, Velculescu V, et al. Gene expression profiles in normal and cancer cells. *Science*. 1997;276:1268-1272.
  15. Morin PJ, Sparks AB, Korinek V, et al. Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. *Science*. 1997;275:1787-1790.
  16. Bahou W, Campbell A, Wicha M. cDNA cloning and molecular characterization of MSE55: a novel human serum constituent protein that displays bone marrow stromal endothelial cell-specific expression. *J Biol Chem*. 1992;267:13986-13992.
  17. Bahou W, Collier B, Potter C, Norton K, Kulok J, Goligorsky M. The thrombin receptor extracellular domain contains sites crucial for peptide ligand-induced activation. *J Clin Invest*. 1993;91:1405-1413.
  18. Heid C, Stevens J, Livak K, Williams P. Real-time quantitative PCR. *Genome Res*. 1996;6:986-994.
  19. Dunn J, McCorkle S, Fraissman L, et al. Genome signature tags (GSTs): a system for profiling genomic DNA. *Nucleic Acid Res*. 2001;29:137-140.
  20. Kroll T, Wolff S. Ranking: a closer look on globalization methods for normalization of gene expression arrays. *Nucleic Acids Res*. 2002;30:e50.
  21. Pruitt KD, Maglott DR. RefSeq and LocusLink: NCBI gene-centered resources. *Nucleic Acids Res*. 2001;29:137-140.
  22. Fox JE, Boyles JK, Reynolds CC, Phillips DR. Actin filament content and organization in unstimulated platelets. *J Cell Biol*. 1984;98:1985-1991.
  23. Safer D, Etzling M, Nachmias VT. Thymosin beta 4 and Fx, an actin-sequestering peptide, are indistinguishable. *J Biol Chem*. 1991;266:4029-4032.
  24. Wallace DC. Mouse models for mitochondrial disease. *Am J Med Genet*. 2001;106:71-93.
  25. Welle S, Bhatt K, Thornton C. Inventory of high-abundance mRNAs in skeletal muscle of normal men. *Genome Res*. 1999;9:506-513.
  26. Karpaluk S, Charizatz A, Langer RM. Glycogenesis and glycconeogenesis in human platelets. Incorporation of glucose, pyruvate, and citrate into platelet glycogen; glycogen synthetase and fructose-1,6-diphosphatase activity. *J Clin Invest*. 1970;49:140-149.
  27. Roha S, Robinson BH. Mitochondria, oxygen free radicals, and apoptosis. *Am J Med Genet*. 2001;106:62-70.
  28. Yu J, Zhang L, Hwang P, Rago C, Kinzler K, Vogelstein B. Identification and classification of p53-regulated genes. *Proc Natl Acad Sci U S A*. 1999;96:14517-14522.
  29. Wang E, Miller L, Ohnmacht G, Liu E, Marincola F. High-fidelity mRNA amplification for gene profiling. *Nat Biotechnol*. 2000;18:157-159.
  30. Kirschbaum L, Sharpe JA, Murphy B, et al. Molecular cloning and characterization of the novel human complement-associated protein, SP-40, 40: a link between the complement and reproductive systems. *EMBO J*. 1989;8:711-718.
  31. McLaughlin L, Zhu G, Mistry M, et al. Apolipoprotein J clusters limits the severity of murine autoimmune myocarditis. *J Clin Invest*. 2000;6:1105-1113.
  32. Martinez DA, Perez JL, Bernal J, Coloma A. Structure, organization, and chromosomal mapping of the human neurogranin gene (NRGN). *Genomics*. 1997;41:243-249.
  33. Wu J, Li J, Huang K, Huang F. Attenuation of PKC and PKA signal transduction in the neurogranin knockout mouse. *J Biol Chem*. 2002;277:19498-19505.
  34. Gygi S, Rochon Y, Franzos B, Aebersold R. Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol*. 1999;19:1720-1730.
  35. Link A, Eng J, Schieltz DM, et al. Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol*. 1999;17:676-682.
  36. Unlu M, Morgan M, Minden J. Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis*. 1997;18:2071-2077.
  37. Gygi S, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol*. 1999;17:994-999.
  38. Halizmanikalis V, Lee K. Dynamical analysis of gene networks requires both mRNA and protein expression information. *Metabol Eng*. 1999;1:275-281.



**Overexpression of a DEAD box protein (DDX1) in neuroblastoma and retinoblastoma cell lines.**

Godbout R, Packer M, Bie W.

Department of Oncology, Cross Cancer Institute and University of Alberta, 11560 University Ave., Edmonton, Alberta T6G1Z2, Canada.

The DEAD box gene, DDX1, is a putative RNA helicase that is co-amplified with MYCN in a subset of retinoblastoma (RB) and neuroblastoma (NB) tumors and cell lines. Although gene amplification usually involves hundreds to thousands of kilobase pairs of DNA, a number of studies suggest that co-amplified genes are only overexpressed if they provide a selective advantage to the cells in which they are amplified. Here, we further characterize DDX1 by identifying its putative transcription and translation initiation sites. We analyze DDX1 protein levels in MYCN/DDX1-amplified NB and RB cell lines using polyclonal antibodies specific to DDX1 and show that there is a good correlation with DDX1 gene copy number, DDX1 transcript levels, and DDX1 protein levels in all cell lines studied. DDX1 protein is found in both the nucleus and cytoplasm of DDX1-amplified lines but is localized primarily to the nucleus of nonamplified cells. Our results indicate that DDX1 may be involved in either the formation or progression of a subset of NB and RB tumors and suggest that DDX1 normally plays a role in the metabolism of RNAs located in the nucleus of the cell.

PMID: 9694872 [PubMed - indexed for MEDLINE]

108: J Endocrinol. 2003 May;177(2):327-35.

Related Articles, Links

Society of Endocrinology  
FREE FULL TEXT

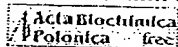
### Modulation of gap junction mediated intercellular communication in TM3 Leydig cells.

Goldenberg RC, Fortes ES, Cristancho JM, Morales MM, Franci CR, Varanda WA, Campos de Carvalho AC.

Institute of Biophysics Carlos Chagas Filho, UFRJ, Brazil.

Long-term modulation of intercellular communication via gap junctions was investigated in TM3 Leydig cells, under low and high confluence states, and upon treatment of the cells for different times with activators of protein kinase A (PKA) and protein kinase C (PKC). Cells in low confluence were readily coupled, as determined by transfer of the dye Lucifer Yellow; on reaching confluence, the cells uncoupled. Western blots and RT-PCR revealed that connexin 43 (Cx43) was abundantly expressed in TM3 Leydig cells and its expression was decreased after the cells achieved confluence. Stimulation of PKA or PKC induced a decrease in cell-cell communication. Staurosporin, an inhibitor of protein kinases, increased coupling and was able to prevent and reverse the uncoupling actions of dibutyryl cAMP and 12-O-tetradecanoyl-phorbol-13-acetate (TPA). Under modulation by confluence, Cx43 was localized to the appositional membranes when cells were coupled and was mainly in the cytoplasm when they were uncoupled. In addition, cAMP and TPA reduced the surface membrane labeling for Cx43, whereas staurosporin increased it. These data show a strong correlation between functional coupling and the membrane distribution of Cx43, implying that this connexin has an important role in intercellular communication between TM3 cells. Furthermore, increased testosterone secretion in response to luteinizing hormone was accompanied by a decrease in intercellular communication, suggesting that gap junction mediated coupling may be a modulator of hormone secretion in TM3 cells.

PMID: 12740021 [PubMed - indexed for MEDLINE]



## Expression level of Ubc9 protein in rat tissues.

Golebiowski F, Szulc A, Sakowicz M, Szutowicz A, Pawelczyk T.

Department of Molecular Medicine, Medical University of Gdansk, 80-211 Gdansk, Poland.

Ubc9 is a homologue of the E2 ubiquitin conjugating enzyme and participates in the covalent linking of SUMO-1 molecule to the target protein. In this report we describe a simple and efficient method for obtaining pure human recombinant Ubc9 protein. The purified Ubc9 retained its native structure and was fully active in an in vitro sumoylation assay with the promyelocytic leukaemia (PML) peptide as a substrate. In order to better understand the physiology of Ubc9 protein we examined its levels in several rat tissues. Immunoblot analyses performed on tissue extracts revealed quantitative and qualitative differences in the expression pattern of Ubc9. The Ubc9 protein was present at a high level in spleen and lung. Moderate level of Ubc9 was detected in kidney and liver. Low amount of Ubc9 was observed in brain, whereas the 18 kDa band of Ubc9 was barely visible or absent in heart and skeletal muscle. In heart and muscle extracts the Ubc9 antibodies recognized a 38 kDa protein band. This band was not visible in extracts of other rat tissues. A comparison of the relative levels of Ubc9 mRNA and protein indicated that the overall expression level of Ubc9 was the highest in spleen and lung. In spleen, lung, kidney, brain, liver and heart there was a good correlation between the 18 kDa protein and Ubc9 mRNA levels. In skeletal muscle the Ubc9 mRNA level was unproportionally high comparing to the level of the 18 kDa protein. The presented data indicate that in the rat the expression of the Ubc9 protein appears to have some degree of tissue specificity.

PMID: 14739995 [PubMed - indexed for MEDLINE]



# Analysis of mRNA expression and protein abundance data: an approach for the comparison of the enrichment of features in the cellular population of proteins and transcripts

Dov Greenbaum<sup>3,†</sup>, Ronald Jansen<sup>1,†</sup> and Mark Gerstein<sup>1,2,\*</sup>

<sup>1</sup>Departments of Molecular Biophysics & Biochemistry, <sup>2</sup>Computer Science and <sup>3</sup>Genetics, 266 Whitney Avenue, Yale University, PO Box 208114, New Haven, CT 06520, USA

Received on July 2, 2001; revised on October 5, 2001; accepted on October 22, 2001

## ABSTRACT

**Motivation:** Protein abundance is related to mRNA expression through many different cellular processes. Up to now, there have been conflicting results on how correlated the levels of these two quantities are. Given that expression and abundance data are significantly more complex and noisy than the underlying genomic sequence information, it is reasonable to simplify and average them in terms of broad proteomic categories and features (e.g. functions or secondary structures), for understanding their relationship. Furthermore, it will be essential to integrate, within a common framework, the results of many varied experiments by different investigators. This will allow one to survey the characteristics of highly expressed genes and proteins.

**Results:** To this end, we outline a formalism for merging and scaling many different gene expression and protein abundance data sets into a comprehensive reference set, and we develop an approach for analyzing this in terms of broad categories, such as composition, function, structure and localization. As the various experiments are not always done using the same set of genes, sampling bias becomes a central issue, and our formalism is designed to explicitly show this and correct for it. We apply our formalism to the currently available gene expression and protein abundance data for yeast. Overall, we found substantial agreement between gene expression and protein abundance, in terms of the enrichment of structural and functional categories. This agreement, which was considerably greater than the simple correlation between these quantities for individual genes, reflects the way broad categories collect many individual measurements into simple, robust averages. In particular, we found

that in comparison to the population of genes in the yeast genome, the cellular populations of transcripts and proteins (weighted by their respective abundances, the transcriptome and what we dub the translome) were both enriched in: (i) the small amino acids Val, Gly, and Ala; (ii) low molecular weight proteins; (iii) helices and sheets relative to coils; (iv) cytoplasmic proteins relative to nuclear ones; and (v) proteins involved in 'protein synthesis,' 'cell structure,' and 'energy production.'

**Supplementary information:** <http://genecensus.org/expression/translatome>

**Contact:** mark.gerstein@yale.edu

## INTRODUCTION

High throughput experimentation, measuring mRNA (Schena *et al.*, 1995; Eisen and Brown, 1999; Feren and Brown, 1999; Lipshutz *et al.*, 1999) and protein expression (Anderson and Seilhamer, 1997; Futcher *et al.*, 1999; Gygi *et al.*, 1999a; Ross-Macdonald *et al.*, 1999; Lopez, 2000; MacBeath and Schreiber, 2000; Nelson *et al.*, 2000; Zhu *et al.*, 2000) are currently the single richest source of genomic information. However, how to best interpret this data is still an open question (Bassett *et al.*, 1996; Wiltes and Friedman, 1999; Zhang, 1999; Gerstein and Jansen, 2000; Searls, 2000; Sherlock, 2000; Claverie, 1999; Einarson and Golemis, 2000; Epstein and Butow, 2000; Shapiro and Harris, 2000). Understanding how protein abundance is related to mRNA transcript levels is essential for interpreting gene expression, protein interactions, structures and functions in a cellular system (Hatzimanikatis *et al.*, 1999). Moreover, as protein concentration is the more relevant variable with respect to enzyme activity, it connects genomics to the physical chemistry of the cell (Kidd *et al.*, 2001). Protein abundance may also be invaluable for diagnostics and for determining drug targets (Corthals *et al.*, 2000).

\*To whom correspondence should be addressed.

†These authors contributed equally to this work.



Previously, we surveyed the population of protein features—such as folds, amino acid composition, and functions—in yeast, and other recently sequenced genomes (Gerstein, 1997, 1998a,b; Gerstein and Hegyi, 1998; Hegyi and Gerstein, 1999; Das and Gerstein, 2000; Lin and Gerstein, 2000), and we extended this concept to compare the population of features in the yeast transcriptome to that in the genome (Drawid *et al.*, 2000; Jansen and Gerstein, 2000). Others have also done related work (Frishman and Mewes, 1997; Tatusov *et al.*, 1997; Jones, 1998; Wallin and von Heijne, 1998; Frishman and Mewes, 1999; Wolf *et al.*, 1999). Here, we present a new methodology to compare the features of the mRNA expression population with the protein abundance population.

Precise terminology is essential for this comparison. Unfortunately, 'proteome' is used inconsistently. Proteome can logically be used to describe all the distinct proteins in the genome (Qi *et al.*, 1996; Cavalcoli *et al.*, 1997; Fey *et al.*, 1997; Garrels *et al.*, 1997; Gaasterland, 1999; Jones, 1999; Sali, 1999; Tekala *et al.*, 1999; Bairoch, 2000; Cambillau and Claverie, 2000; Doolittle, 2000; Pandey and Mann, 2000; Rubin *et al.*, 2000) and, in this context, it is equivalent to what others may refer to as the coding part of the genome. However, in papers on two-dimensional (2D) electrophoresis, it is often used to describe the sum total of proteins in a cell, taking into account the different levels of protein abundance (Shevchenko *et al.*, 1996; Gygi *et al.*, 2000a; Lopez, 2000; Washburn and Yates, 2000). In an effort to be clear, we propose the term 'translatome' for this second usage of proteome.

With this definition, we are able to refer compactly to three different cellular populations. These are illustrated in Figure 1.

- (i) We use the term *genome* when we refer to the population of open reading frames, where each ORF counts once.
- (ii) We use the term *transcriptome* when we refer to the population of mRNA transcripts. This term was originally coined by Velculescu *et al.* (1997). Note that each ORF may give rise to different numbers of transcripts. Consequently, the transcriptome is essentially the same as the genome but with each ORF weighted by its expression level.
- (iii) The next level is the cellular population of proteins. As each protein represents a translated transcript, we make an analogy with the term transcriptome and use the term *translatome* as described above to describe this third population. Thus, the translatome is a subset of the genome where each ORF is weighted by its associated level of protein abundance.

Note that one could also, less compactly call the translatome a 'weighted proteome.' However, doing so assumes one of the two aforementioned definitions of proteome. To avoid ambiguity, we studiously avoid the use of proteome altogether in the paper.

Differences between the translatome and the transcriptome exist given that transcripts from different genes can give rise to different numbers of proteins, due to different rates of translation and protein degradation. Post-transcriptional modifications further affect the translatome.

In our analysis of the transcriptome and translatome, we focus on global protein features rather than the comparison of individual genes. Previous analyses have shown that differences between mRNA expression and protein abundance levels can be quite dramatic for individual genes. This may either be due to the noise in the data or to fundamental biological processes. However, our analyses show that the variation between transcriptome and translatome is much smaller for global properties that are computed by averaging over the properties of many individual genes.

## METHODS

### Data sources used

For our analysis we culled many divergent data sets, representing protein abundance and mRNA expression experiments and also other sources of genome annotation. These are all summarized in Table 1.

### Biases in the data

The databases that annotate the specific genes may not always be accurate (Ishii *et al.*, 2000). Gene Chip experiments suffer with regard to cross hybridization and the saturation of probes. SAGE data degrades for lowly expressed mRNAs. 2D gels are unable to resolve membrane proteins (approximately 30% of the genome) and basic proteins (Gerstein, 1998c; Krögh *et al.*, 2001). In addition, the procedures for identification and quantification of the protein spots are subject to uncertainties (Haynes and Yates, 2000). Human biases include the lack of low abundance proteins (Fey and Larsen, 2001; Gygi *et al.*, 2000b; Harry *et al.*, 2000) and the differences between laboratories in sample preparation. Our reference expression data set attempts to resolve these problems.

### Data set scaling

*A reference set for mRNA expression.* With many different mRNA expression data sets available, it is worthwhile to integrate them into a single unified reference set, with the intention of reducing the noise and errors contained in the individual data sets and to obtain a unified estimate of the normal expression state in a cell.

We adopt an iterative scaling and merging formalism,

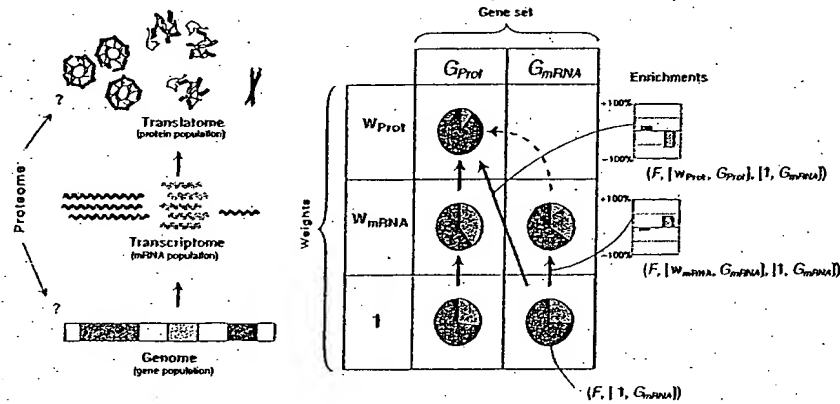


Fig. 1. Schematic overview of the analysis. On the left-side we outline the terms we use to describe the process of gene expression. The coding section of the genome is transcribed into a population of mRNA transcripts called the 'transcriptome.' The transcripts in turn are translated to a population of proteins; we use the term 'translatome' for this protein population rather than the alternative 'proteome' because the latter term may be confounded with the protein complement of the genome (which is not necessarily associated with a quantitative abundance level).

The matrix in the middle schematically shows an analysis of the three stages of expression. In general, we define a protein 'population' as a set of genes associated with a corresponding number of expression or abundance levels ('weights'). In the matrix each row represents a weight and each column a gene set. In particular, we differentiate between the mRNA reference expression set ( $G_{mRNA} = G_{Gen}$ ), which essentially covers the complete genome, and the reference protein abundance set ( $G_{Prot}$ ) which contains the proteins in data sets 2-DE #1 and 2-DE #2 (see Table 1) because the protein abundance set is a significantly smaller subset of the genome. By definition, this subset contains only proteins that can be identified by 2-D gel electrophoresis and is therefore biased in this sense. The enrichment figures throughout this paper, through a comparison of the right- and left-sides of this figure, show the results of the experimental biases of 2D gels on the data set. Each pie chart represents a composition of a particular protein feature  $F$  (for instance, an amino acid composition) in a population (represented by the symbol  $\mu$ ). We can further look at the 'enrichment' of this feature in one population relative to another (represented by the symbol  $\Delta$ , see Section 'Methods' for an explanation of the formalism).

which we summarize below. We present a more detailed review of the methods on our web site.

We start with the values of one gene chip data set  $U_i$  where  $i$  is used throughout as a subscript to denote gene number. We then transform the values of the next Gene Chip data set  $X_i$  to  $Y_i$  with the following non-linear regression:  $\min \sum_i (Y_i - U_i)^2$  with  $Y_i = AX_i^B$  where  $A$  and  $B$  are the parameters of the regression. Note that two Gene Chip sets may not be defined for the same set of genes, so we have to perform the fit only over the genes common to both sets. The motivation for scaling is that the dynamic range of observed expression levels varies somewhat between different data sets, although cell types and growth conditions are very similar. Reasons for disparity may include different calibration procedures for relating fluorescence intensity to a cellular concentration (measured in copies of transcripts per cell) or different protocols for harvesting and reverse-transcribing the cellular mRNA.

We then merge and average the data to create a new

reference set  $V$  as follows:

$$\text{If } U_i \text{ and } Y_i \text{ are both defined for gene } i \text{ and } \frac{|Y_i - U_i|}{Y_i + U_i} < \alpha$$

$$\text{Then } V_i = \frac{1}{2}(Y_i + U_i)$$

$$\text{Else if only } Y_i \text{ exists, } V_i = Y_i$$

$$\text{Else } V_i = U_i.$$

As presented above, where only one data set has a value for the corresponding ORF, we incorporated that value and did not exclude it. When both data sets have values for an ORF, we averaged the values if they were within 15% of each other; otherwise, we just stayed with the original chip data set  $U_i$ . We used  $\alpha = 15\%$  in order to prevent outliers from skewing the result. This 15% value is a reasonable threshold for excluding outliers though other values (e.g. 10 or 20%) would give similar results (data not shown). Other data sets are subsequently included in the same procedure, continuing the iteration from the new

expression values  $V_i$ . The initial iteration starts with the Young Expression Set, as  $U_i$ , since we have the highest confidence in its accuracy.

The SAGE data (Velculescu *et al.*, 1997) was not included in the above procedure since it is of a fundamentally different nature. An advantage of the SAGE technology over Gene Chips is that there is no possible signal saturation for high expression levels, as is possible for chips (Futcher *et al.*, 1999). Conversely, SAGE values are less reliable for lowly expressed genes since there is a chance that one might not sequence a SAGE tag corresponding to such a gene altogether. Therefore, if after the last iteration, the average Gene Chip expression level  $V_i$  was both above a certain threshold  $\beta$  and below the SAGE expression level  $S_i$  for the same gene, it was replaced with the SAGE value; otherwise the average Gene Chip value was kept. This gave us our final expression set  $w_{\text{mRNA}}$ . Our treatment of the SAGE data is modeled after that in Futcher *et al.* (1999), and like them, we used  $\beta = 16$ .

This incorporation of the SAGE data into the reference data set ensures that the highly expressed outliers are as accurate as possible.

Rather than plain arithmetic averaging, this overall scaling procedure with the  $\alpha$  cutoff avoids 'artificial averages' that combine very different values for a particular gene. Some expression values might be statistical outliers. In addition, it may be possible that the expression levels of a variety of genes can only be within mutually exclusive ranges or modes, such as when two alternative pathways are switched on or off. Simply averaging these would give values that are less representative of the particular mode values. This situation is analogous to that in averaging together an ensemble of protein structures (i.e. from NMR structure determination). Each structure could be stereochemically correct, with all side-chain atoms in predefined rotamer configurations. However, an average of all structures could yield one that is stereochemically incorrect if this involved averaging over particular side-chains in different rotameric states.

With regard to our regression analysis, we have investigated both non-linear and linear fits but found a non-linear procedure to be more advantageous. The non-linear relationship between different expression data sets perhaps reflects saturation in one or more of the Gene Chips—not an uncommon phenomenon. This non-linearity is immediately evident on scatter plots of two data sets against one another (see website). Accordingly, the non-linear fit produces a smaller residual than the linear fit: 98 297 (non-linear) versus 122 182 (linear) for the scaling of the Church data set and 59 828 (non-linear) versus 67 462 (linear) for the Samson data set.

**A reference set for protein abundance.** We followed a similar procedure to calculate a reference protein abundance set from the two gel electrophoresis data sets. We first scaled the two data sets against the mRNA expression reference data set, getting regression parameters  $C_j$  and  $D_j$ :

$$\min \sum_i (P_{i,j} - C_j w_{\text{mRNA},i}^{D_j})^2$$

where the subscript  $j$  indicates the data set 2-DE #1 or 2-DE #2 respectively;  $P_{i,j}$  is the protein abundance value in data set  $j$ , and  $w_{\text{mRNA},i}$  the corresponding reference expression value, and  $C_j$  and  $D_j$  are the parameters of the non-linear regression.

Using these parameters, we transformed the values of set 2-DE #2 onto 2-DE #1. Then we combined both sets into the reference protein set  $w_{\text{Prot}}$  by averaging them, if both values existed. Otherwise, by using the existing value, viz:

$$Q_{i,2} \equiv C_1 \left( \frac{P_{i,2}}{C_2} \right)^{D_1/D_2}$$

$w_{\text{Prot},i} = (P_{i,1} + Q_{i,2})/2$  if both  $P_{i,1}$  and  $Q_{i,2}$  exist.

Else if only  $P_{i,1}$  exists,  $w_{\text{Prot},i} = P_{i,1}$

Else if  $Q_{i,2}$  exists,  $w_{\text{Prot},i} = Q_{i,2}$ .

#### Enrichment of features

**Formalism.** In the next part of our analysis, we want to group a number of proteins together into various categories based on common features and characterize those features that are enriched in one population relative to another, i.e. the translome population of proteins as measured by 2D gels relative to the transcriptome population of transcripts or the genome population of genes. To this end, we set up a formalism that could be applied universally to all the attributes that we were interested in. Due to the limitations of the experiments, the translome, transcriptome, and genome populations are defined on different sets of genes, and sometimes we want to remove this 'selection bias' by forcing them to be compared on exactly the same set of genes. This is a key aspect of our formalism as presented in Figure 1.

We call an entity like  $\{w, G\}$  a 'population,' where  $G$  is a set describing a particular selection of genes from the genome and  $w$  is vector of weights associated with each element of this population. In particular, we focus on three main populations here:

- (i)  $\{1, G_{\text{Gen}}\}$  is the population of genes in the genome, all 6280 genes weighted once ( $w = 1$ );
- (ii)  $\{w_{\text{mRNA}}, G_{\text{mRNA}}\}$  is the observed population of the transcripts in the transcriptome, i.e. the 6249 genes in the reference expression set weighted by their reference expression value;

- (iii)  $[w_{\text{Prot}}, G_{\text{Prot}}]$  is the observed cellular population of the proteins in the translome, i.e. the 181 genes in the reference abundance set weighted by their reference abundance value.

(The set of genes in the genome  $G_{\text{Gen}}$  is approximately equal to the genes in set  $G_{\text{mRNA}}$ , such that we can use both symbols interchangeably.) We can also use this notation to describe specific experiments—e.g.  $[w_{\text{lacZ}}, G_{\text{lacZ}}]$  describes the gene set and weights relating to the transposon abundance set.

Furthermore, we define  $F_j$  as the value of a feature  $F$  in ORF  $j$ . For example,  $F$  could be the composition of leucine (a real number) or a binary value (0 or 1) indicating whether an ORF contains a trans-membrane segment. Given these definitions, the weighted average of feature  $F$  in population  $[w, G]$  is:

$$\mu(F, [w, G]) = \frac{\sum_{j \in G} w_j F_j}{\sum_{j \in G} w_j}$$

The weighted averages of two populations  $[w, G]$  and  $[v, S]$  can be compared by simply looking at their relative difference  $\Delta$ :

$$\Delta(F, [v, S], [w, G]) = \frac{\mu(F, [v, S]) - \mu(F, [w, G])}{\mu(F, [w, G])}$$

where  $v$  and  $w$  are weights for the sets of ORFs  $S$  and  $G$  respectively. We call  $\Delta$  the 'enrichment' of feature  $F$  because it indicates whether  $F$  is enriched (if  $\Delta$  is positive) or depleted (if  $\Delta$  is negative) in population  $[v, S]$  relative to  $[w, G]$ .

Usually, the gene set  $G$  is defined by the particular experiment, for which the weight  $w$  was measured. However, it is also possible to combine the gene set associated with one experiment with expression levels from another set. One may want to do this to compute the enrichment only on the genes common to both populations, for which there are defined values for both  $w$  and  $v$ , viz:  $\Delta(F, [v, S \cap G], [w, S \cap G])$ . In practice, this is most relevant for comparing  $G_{\text{Prot}}$  and  $G_{\text{mRNA}}$ . Since  $G_{\text{Prot}}$  is completely a subset of  $G_{\text{mRNA}}$ , we need not explicitly deal with intersections if we calculate all statistics directly over  $G_{\text{Prot}}$ .

One can adjust the weight vectors to take into account different types of averaging. For instance, when computing the amino acid composition ( $F = aa$ ) from the amino acid compositions of individual ORFs  $F_j = aa_j$  ( $\forall j \in G$ ), we weight by ORF length. In the case of expression weights, we have:

$$w_j = N_j w_{\text{mRNA}, j} \quad \forall j \in G$$

where  $N_j$  is a measure of the length of ORF  $j$  (such as the number of amino acids).

On the other hand, when computing the average molecular weight per amino acid, we need to normalize by the number of amino acids per ORF, which is equivalent to choosing the following weights:

$$w_j = \frac{w_{\text{mRNA}, j}}{N_j} \quad \forall j \in G.$$

#### Application of methodology to quantitative abundance sets

Having defined our formalism, we applied it to a diverse set of protein features in yeast.

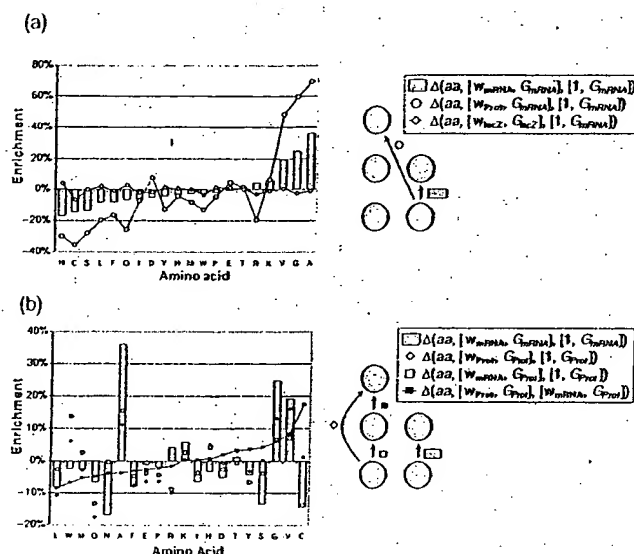
**Amino acid enrichment.** As shown in Figure 2a, we used our methodology to measure the enrichment of individual amino acids in both the translome and the transcriptome relative to the genome. We found that three amino acids—valine, glycine and alanine—were consistently enriched in both transcriptome and translome populations.

In Figure 2a we compare different gene sets. In Figure 2b we focus mainly on the variation in enrichments when all the comparisons are restricted to the set of 181 genes ( $G_{\text{Prot}} \cap G_{\text{mRNA}} = G_{\text{Prot}}$ ) common to all data sets. Thus, the differences between the populations now only reflect the effects of differential transcription of certain genes and differential translation of certain transcripts. We find here an enrichment specifically of cysteine in the translome in relation to the transcriptome.

To measure the statistical significance of the results on amino acid enrichment, we have performed a control analysis on a randomized data set (Figure 2d). We randomly permuted the expression values of the ORFs 1000 times and then recomputed the enrichments. This allowed us to compute distributions for the amino acid enrichments and, from integrating these, one-sided  $p$ -values indicating the significance of the observed enrichments.

**Amino acid enrichment in Transposon data set.** We also tried to extend our methodology, ineffectively, to cope with the semi-quantitative Transposon set. We used only those 450 ORFs that consistently yielded either no expression or high expression, as binary data, on or off. We show the enrichments of amino acids computed from this filtered Transposon abundance set in Figure 2a. Overall, the enrichments from this set seemed to be attenuated in comparison to other data.

**Biomass enrichment.** A corollary to amino acid enrichments is the determination of the average biomass of the transcriptome and translome populations (shown in Figure 2c). We found that the average molecular weight of a protein in both populations was, on average, lower than in the genome population. These preliminary observations suggest a cell preference to use less energetically expensive proteins for those that are highly-transcribed or trans-



**Fig. 2.** Amino acid and biomass enrichment. (a) Shows the amino acid enrichments between different populations as indicated by the legend to the right of the plot (the legend is ordered in the same way as the schematic illustration in Figure 1). The bars indicate the enrichment of the transcriptome relative to the genome, whereas the circles indicate the enrichment of the translatome relative to the genome. In addition, we also show the enrichment for protein abundance from the Transposon abundance set, represented by the circles with the line through them. (b) Shows a different view of amino acid enrichment from that contained in (a), now focusing on changes, and thus restricting the comparison to the genes common to all the data sets. The graph is ordered according to the enrichment from transcriptome to translatome (black squares). We focus here only on the changes for the abundance gene set ( $G_{Prot}$ ) to exclude the effects that arise from looking at different subsets. In this view the enrichments from genome to transcriptome (white squares) and from genome to translatome (white diamonds) look more similar than do the analogous sets in (a). To make comparison with (a) easier we again show the enrichment from genome to the transcriptome for the complete gene set ( $G_{Gen}$ , shown in bars). (c) Shows biomass enrichment. The left panel depicts the average molecular weight per ORF (in units of kDa) and the right panel, the average molecular weight per amino acid (in units of Daltons) in each of the three stages of gene expression. The numbers inside the circles indicate the average molecular weights. The values next to the arrows indicate the enrichments in biomass between different populations. Both the circle diameters and the arrow widths are functions of the corresponding values (the hollow arrow indicates a positive value). It is very clear that the average molecular weight per ORF is much lower in the translatome (by 20 or 15%) and transcriptome (by 29%) than in the genome. This relative depletion of biomass mainly takes place as a result of transcription; the effect of translation is less clear, depending on the populations compared. On the other hand, the depletion in the average molecular weight per amino acid ( $-3.3\%$  from genome to translatome) is an order of magnitude smaller than in the average weight per ORF. This shows that the yeast cell favors the expression of shorter ORFs over longer ones, and agrees with our earlier observation that there is a negative correlation between maximum ORF length and mRNA expression (Jansen and Gerstein, 2000); it seems that this effect mainly takes place during transcription rather than translation. (d) This plot shows that the amino acid enrichments are statistically significant. We have assessed significance by randomly permuting the expression levels among the genes and then recomputing the amino acid enrichments. This procedure can be repeated and used to generate distributions of random enrichments that can then be compared against the observed enrichments. In the plot the gray bars represent the observed enrichments already shown in Figure 3a. On top of the gray bars we show standard boxplots of enrichment distributions based on 1000 random permutations. (The middle line represents the distribution median. The upper and lower sides of the box coincide with the upper and lower quantiles. Outliers are shown as dots and defined as data points that are outside the range of the whiskers, the length of which is 1.5 the interquartile distance.) Based on the random distributions, we can compute one-sided  $p$ -values for the observed enrichments. Amino acids for which the  $p$ -values are less than  $10^{-3}$  are shown in bold font.

lated. However, we also found that the average molecular weight per amino acid differed much less between the transcriptome and the translatome on the one hand, and the

genome on the other hand (though it was still slightly less). This finding indicates that lower molecular weights in the translatome and transcriptome relative to the genome are

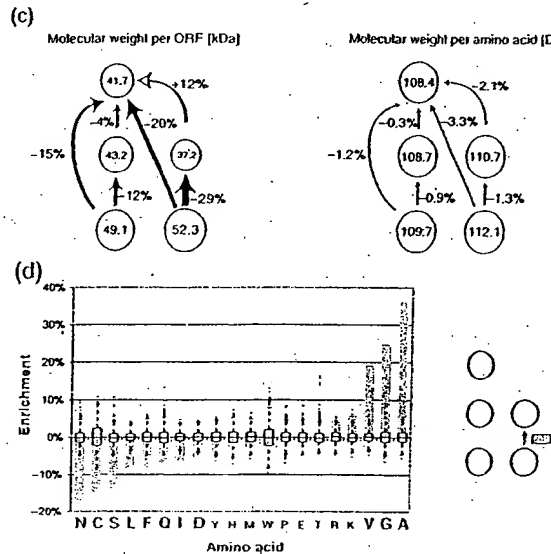


Fig. 2. Cont.

predominantly due to greater expression of shorter proteins rather than the incorporation of smaller amino acids.

**Secondary structure composition.** We also used our methodology to study the enrichment of secondary-structural features. Secondary structural annotation was derived from structure prediction applied uniformly to all the ORFs in the yeast genome as described in Table 1. As shown in Figure 3a, all three populations—genome, transcriptome, and translatome—had a fairly similar composition of secondary structures—sheets, helices, and coils. The differences between populations were marginal and based only on the small subset of genes.

We also found that Transmembrane (TM) proteins were significantly depleted in the transcriptome (see website and caption). These results are consistent with our previous analyses (Jansen and Gerstein, 2000). The protein abundance data does not have any membrane proteins.

**Subcellular localization.** Figure 3c shows the enrichment of proteins associated with the various subcellular compartments. For clarity, we divided the cell into five distinct subcellular compartments, (see Table 1). We found that, in comparison to the genome, both the transcriptome and translatome are enriched in cytoplasmic proteins. This is true whether we make our comparisons in

relation to the relatively large reference mRNA expression set or the smaller reference protein abundance set. As Figure 3c shows, the 2D gel experiments are clearly biased towards proteins from the cytoplasm. However, in the biased subset  $G_{P101}$  transcription and translation lead to an even higher fraction of cytoplasmic proteins in the translatome.

**Functional categories.** Finally, we compared the enrichment of various functional categories in both the translatome and the transcriptome (see Figure 3b). This gives us a broad yet informative view of the cell as a whole. As described in Table 1, we used the top-level of the MIPS scheme for the functional category definitions. We found broad differences between the various populations, with some of the functional categories showing strikingly high enrichments.

## DISCUSSION AND CONCLUSION

We developed: (i) a methodology for integrating many different types of gene expression and protein abundance into a common framework and applied this to a preliminary analysis; (ii) a procedure for scaling and merging different mRNA and protein sets together; and (iii) an approach for computing the enrichment of various proteomic features in the population of transcripts and proteins. We showed that by analyzing broad categories instead of individual noisy

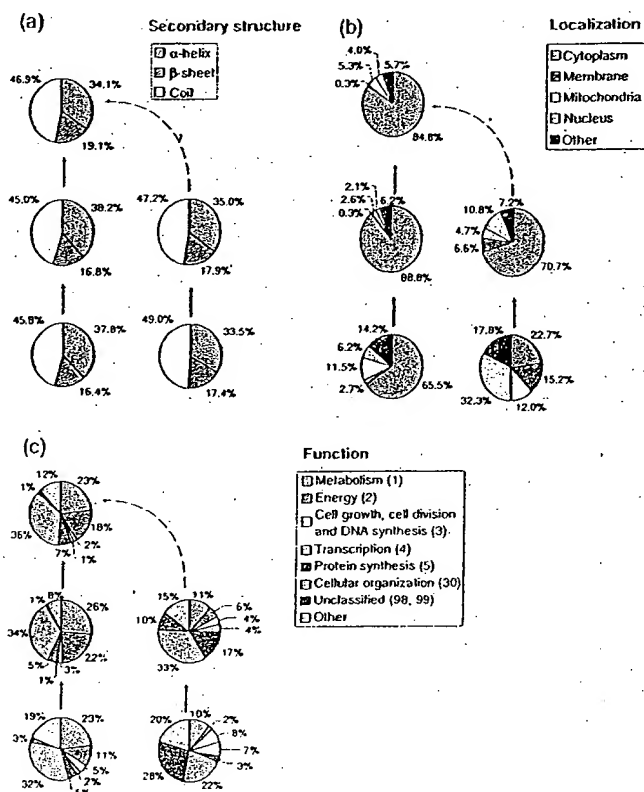


Fig. 3. Breakdown of the transcriptome and translome in terms of broad categories relating to structure, localization, and function. All of the subfigures are analogous to the schematic illustration in Figure 1. (a) Represents the composition of secondary structure in the different populations. (b) Represents the distribution of subcellular localizations associated with proteins in the various populations. We used standardized localizations developed earlier (Drawid and Geisstein, 2000), which, in turn, were derived from the MIPS, YPD, and SwissProt databases (Bairoch and Apweiler, 2000; Costanzo *et al.*, 2000; Mewes *et al.*, 2000). The subcellular localization has been experimentally determined for less than half of the yeast proteins, so our analysis applies only to this subset. (c) Shows the division of ORFs into different functional categories (according to the MIPS classification) in the various populations. Only the largest functional categories of the top level of the MIPS classification are shown. The group 'other' contains the smaller top-level categories lumped together. This 'other' group is different from the group 'unclassified,' which contains genes without any functional description.

data points, we could find logical trends in the underlying data. For example, individual transcription factors might have higher or lower protein abundance than one expects from their mRNA expression, but the category 'transcription factors' as a whole has a similar representation in the transcriptome and translome.

We found, as previously described (Furber *et al.*, 1999; Gygi *et al.*, 1999b; Greenbaum *et al.*, 2001), a weak correlation between individual measurements of mRNA

and protein abundance. The outliers of this correlation tend to be associated with cellular organization. One might conceive of using these outliers (i.e. those with significantly different transcriptional and translational behavior) to find consensus regulatory sequences. One possible method would involve using predicted mRNA structures (Jaeger *et al.*, 1990; Zuker, 2000) to find and investigate consensus structural elements in these outliers to which the yeast translational machinery is known to be

Table 1. Data sets

Data set	Description	Size [ORFs]	Reference
<b>mRNA expression</b>			
Young	Gene chip profiles yeast cells with mutations that affect transcription	5455	Holstege <i>et al.</i> (1998)
Church	Gene chip profiles of yeast cells under four different conditions	6263	Roth <i>et al.</i> (1998)
Samson	Comparing gene chip profiles for yeast cells subjected to alkylating agent	6090	Jelinsky and Samson (1999)
SAGE	Yeast cells during vegetative growth	3778	Velculescu <i>et al.</i> (1997)
Reference expression	Scaling and integrating the mRNA expression set into one data source	6249	—
<b>Protein abundance</b>			
2-DE #1	Measurement of yeast protein abundance by 2D gel electrophoresis and mass spectrometry	156	Gygi <i>et al.</i> (1999a,b)
2-DE #2	Similar to 2-DE set #1	71	Futcher <i>et al.</i> (1999)
Transposon	Large-scale fusions of yeast genes with <i>lacZ</i> by transposon insertion	1410	Ross-Macdonald <i>et al.</i> (1999)
Reference abundance	Scaling and integrating the 2-DE data sets into one data source	181	—
<b>Annotation</b>			
Annotated localization	Subcellular localizations of yeast proteins	2133 (6280)	Drawid and Gerstein (2000)
TM segments	Predicted TM and soluble proteins in yeast	2710 (6280)	Gerstein (1998a,b,c)
MIPS functions	Functional categories for yeast ORFs	3519 (6194)	Mewes <i>et al.</i> (2000)
GOR secondary structure	Predicted secondary structure yeast ORFs	6280	Gerstein (1998a,b,c)

This table provides an overview of the data sets used in our analysis. The table is divided into three sections. The top section lists different mRNA expression sets. The middle section shows the protein abundance data sets used. The bottom section contains different annotations of protein features. The column 'Data set' lists a shorthand reference to each data set used throughout this paper. The next columns contain a brief description of the data sets, the number of ORFs contained in each of them, and the literature reference. In contrast to the other data we investigated, the reference expression and abundance data sets have been calculated for the purpose of our analysis (see text). An expanded version of the table is available on our web site.

Some further information on the genome annotations:

**Localization.** Protein localization information from YPD, MIPS and SwissProt were merged, filtered and standardized (Bairoch and Apweiler, 2000; Costanzo *et al.*, 2000; Mewes *et al.*, 2000) into five simplified compartments—cytoplasm, nucleus, membrane, extracellular (including proteins in ER and gel), and mitochondrial—according to the protocol in Drawid *et al.* (2000). This yielded a standardized annotation of protein subcellular localization for 2133 out of 6280 ORFs.

**TM segments.** In 2710 out of 6280 yeast ORFs TM segments are predicted to occur, ranging from low to high confidence (732 ORFs). The TM prediction was performed as follows: the values from the scale for amino acids in a window of size 20 (the typical size of a TM helix) were averaged and then compared against a cutoff of  $-1 \text{ kcal mol}^{-1}$ . A value under this cutoff was taken to indicate the existence of a TM helix. Initial hydrophobic stretches corresponding to signal sequences for membrane insertion were excluded. (These have the pattern of a charged residue within the first seven, followed by a stretch of 14 with an average hydrophobicity under the cutoff.) These parameters have been used, tested, and refined on surveys of membrane protein in genomes. 'Sure' membrane proteins had at least two TM-segments with an average hydrophobicity less than  $-2 \text{ kcal mol}^{-1}$  (Rest *et al.*, 1995; Gerstein *et al.*, 2000; Santoni *et al.*, 2000; Senes *et al.*, 2000).

**Functions.** MIPS functional categories have been assigned to 3519 out of 6194 ORFs. (The remainder are assigned to category '98' or '99', which corresponds to unclassified function.)

sensitive (McCarthy, 1998).

In relation to functional categories, we found three trends that were particularly notable: (i) the 'cellular

organization,' 'protein synthesis,' and 'energy production' categories were increasingly enriched as we moved from genome to transcriptome to proteome. In the transcrip-



tome and translome population relative to the genome; (ii) proteins with 'unclassified function' are significantly depleted, perhaps reflecting a bias against studying them; (iii) proteins in the 'transcription' and 'cell growth, cell division, and DNA synthesis' categories were consistently depleted. This reflects the fact that many of these proteins, such as transcription factors, act as 'switches' such that only small quantities of the protein are necessary to activate or deactivate a process. These results concur with previous calculations (Jansen and Gerstein, 2000) wherein we found the transcriptome is enriched specifically with proteins involved in protein synthesis and energy.

#### Limitations given the small size of the protein abundance data

Even with the extended coverage made possible by merging many data sets together into reference sets, the analysis is still limited by the minimal data. This was most applicable to the protein abundance measurements, potentially biasing our statistical results towards certain protein families. Moreover, the 181 proteins in  $G_{\text{Prot}}$  do not represent a random sample. They are skewed towards highly expressed, well-studied proteins. Our methodology attempts to control for this gene-selection bias through our enrichment formalism, which allows one to rather precisely gauge various aspects of the bias. Conversely, many protein features in both the translome and the transcriptome are dominated by highly expressed proteins. Under these circumstances, it is often sufficient to look at this smaller number of dominating proteins to characterize the whole population. This is similar to the development of the codon adaptation index for yeast (Sharp and Li, 1987). While based on only 24 highly expressed proteins, it has proven to be robust in predicting expression levels for the entire genome.

We believe that the essential formalism and approach that we develop will remain quite relevant for future data sets (Smith, 2000).

#### ACKNOWLEDGEMENT

M.G. thanks the Keck foundation for support.

#### REFERENCES

- An, H., Scopes, R.K. et al. (1991) Gel electrophoretic analysis of *Zymomonas mobilis* glycolytic and fermentative enzymes: identification of alcohol dehydrogenase II as a stress protein. *J. Bacteriol.*, **173**, 5975–5982.
- Anderson, L. and Seilhamer, J. (1997) A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis*, **18**, 533–537.
- Bairoch, A. (2000) Serendipity in bioinformatics, the tribulations of a Swiss bioinformatician through exciting times! *Bioinformatics*, **16**, 45–64.
- Bairoch, A. and Apweiler, R. (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.*, **28**, 45–48.
- Bassett, D.E. Jr., Basrai, M.A. et al. (1996) Exploiting the complete yeast genome sequence. *Curr. Opin. Genet. Dev.*, **6**, 763–766.
- Batke, J., Benito, V.A. et al. (1992) A possible *in vivo* mechanism of intermediate transfer by glycolytic enzyme complexes: steady state fluorescence anisotropy analysis of an enzyme complex formation. *Arch. Biochem. Biophys.*, **296**, 654–659.
- Cambillau, C. and Claverie, J.M. (2000) Structural and genomic correlates of hyperthermostability. *J. Biol. Chem.*, **275**, 32383–32386.
- Cavalcoli, J.D., VanBogelen, R.A. et al. (1997) Unique identification of proteins from small genome organisms: theoretical feasibility of high throughput proteome analysis. *Electrophoresis*, **18**, 2703–2708.
- Claverie, J.M. (1999) Computational methods for the identification of differential and coordinated gene expression [in process citation]. *Hum. Mol. Genet.*, **8**, 1821–1832.
- Corthals, G., Wasinger, V.C., Hochstrasser, D.F. and Sanchez, J.C. (2000) The dynamic range of protein expression: a challenge for proteomic research. *Electrophoresis*, **21**, 1104–1115.
- Costanzo, M.C., Hogan, J.D. et al. (2000) The Yeast Proteome Database (YPD) and *Caenorhabditis elegans* Proteome Database (WormPD): comprehensive resources for the organization and comparison of model organism protein information. *Nucleic Acids Res.*, **28**, 73–76.
- Das, R. and Gerstein, M. (2000) The stability of thermophilic proteins: a study based on comprehensive genome comparison. *Funct. Int. Genom.*, **1**, 33–45.
- Doolittle, W.F. (2000) The nature of the universal ancestor and the evolution of the proteome. *Curr. Opin. Struct. Biol.*, **10**, 355–358.
- Drawid, A. and Gerstein, M. (2000) A Bayesian system integrating expression data with sequence patterns for localizing proteins: comprehensive application to the yeast genome. *J. Mol. Biol.*, **301**, 1059–1075.
- Drawid, A., Jansen, R. et al. (2000) Gene expression levels are correlated with protein subcellular localization. *Trends Genet.*, **10**, 426–430.
- Einarsson, M. and Golemis, E. (2000) Encroaching genomics: adapting large-scale science to small academic laboratories. *Physiol. Genom.*, **2**, 85–92.
- Eisen, M.B. and Brown, P.O. (1999) DNA arrays for analysis of gene expression. *Meth. Enzymol.*, **303**, 179–205.
- Epstein, C. and Butow, R. (2000) Microarray technology—enhanced versatility, persistent challenge. *Curr. Opin. Biotechnol.*, **11**, 36–41.
- Ferea, T. and Brown, P. (1999) Observing the living genome. *Curr. Opin. Genet. Dev.*, **9**, 715–722.
- Fey, S.J., Nawrocki, A. et al. (1997) Proteome analysis of *Saccharomyces cerevisiae*: a methodological outline. *Electrophoresis*, **18**, 1361–72.
- Fey, S.J. and Larsen, P.M. (2001) 2D or not 2D. Two-dimensional gel electrophoresis. *Curr. Opin. Chem. Biol.*, **5**, 26–33.
- Frishman, D. and Mewes, H.W. (1997) Protein structural classes in five complete genomes [letter]. *Nat. Struct. Biol.*, **4**, 626–628.
- Frishman, D. and Mewes, H.W. (1999) Genome-based structural biology. *Prog. Biophys. Mol. Biol.*, **72**, 1–17.

- Futcher, B., Latter, G. *et al.* (1999) A sampling of the yeast proteome. *Mol. Cell Biol.*, **19**, 7357–7368.
- Gaasterland, T. (1999) Archaeal genomics. *Curr. Opin. Microbiol.*, **2**, 542–547.
- Garrels, J.I., McLaughlin, C.S. *et al.* (1997) Proteome studies of *Saccharomyces cerevisiae*: identification and characterization of abundant proteins. *Electrophoresis*, **18**, 1347–1360.
- Gerstein, M. (1997) A structural census of genomes: comparing bacterial, eukaryotic, and archaeal genomes in terms of protein structure. *J. Mol. Biol.*, **274**, 562–576.
- Gerstein, M. (1998a) How representative are the known structures of the proteins in a complete genome? A comprehensive structural census. *Fold. Des.*, **3**, 497–512.
- Gerstein, M. (1998b) Patterns of protein-fold usage in eight microbial genomes: a comprehensive structural census. *Proteins*, **33**, 518–534.
- Gerstein, M. (1998c) Patterns of protein-fold usage in eight microbial genomes: a comprehensive structural census. *Proteins*, **33**, 518–534.
- Gerstein, M. and Hegyi, H. (1998) Comparing genomes in terms of protein structure: surveys of a finite parts list. *FEMS Microbiol. Rev.*, **22**, 277–304.
- Gerstein, M. and Jansen, R. (2000) The current excitement in bioinformatics, analysis of whole-genome expression data: how does it relate to protein structure and function. *Curr. Opin. Struct. Biol.*, **10**, 574–584.
- Gerstein, M., Lin, J. *et al.* (2000) Protein folds in the worm genome. *Proc. Symp. Biocomput.*, 30–41.
- Greenbaum, D., Luscombe, N. *et al.* (2001) Interrelating different types of genomic data, from proteome to secretome: coming in on function. *Genome Res.*, **11**, 1463–1468.
- Gygi, S.P., Rist, B. *et al.* (1999a) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnol.*, **17**, 994–999.
- Gygi, S.P., Rochon, Y. *et al.* (1999b) Correlation between protein and mRNA abundance in yeast. *Mol. Cell Biol.*, **19**, 1720–1730.
- Gygi, S.P., Conhals, G.L. *et al.* (2000a) Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc. Natl Acad. Sci. USA*, **97**, 9390–9395.
- Gygi, S.P., Rist, B. *et al.* (2000b) Measuring gene expression by quantitative proteome analysis. *Curr. Opin. Biotechnol.*, **11**, 396–401.
- Harry, J.L., Wilkins, M.R. *et al.* (2000) Proteomics: capacity versus utility. *Electrophoresis*, **21**, 1071–1081.
- Hatzimanikatis, V., Choe, L.H. *et al.* (1999) Proteomics: theoretical and experimental considerations. *Biotechnol. Prog.*, **15**, 312–318.
- Haynes, P.A. and Yates, J.R. (2000) Proteome profiling-pitfalls and progress. *Yeast*, **17**, 81–87.
- Hegyi, H. and Gerstein, M. (1999) The relationship between protein structure and function: a comprehensive survey with application to the yeast genome. *J. Mol. Biol.*, **288**, 147–164.
- Holstege, F.C., Jennings, E.G. *et al.* (1998) Dissecting the regulatory circuitry of a eukaryotic genome. *Cell*, **95**, 717–728.
- Ishii, M., Hashimoto, S. *et al.* (2000) Direct comparison of genechip and SAGE on the quantitative accuracy in transcript profiling analysis. *Genomics*, **68**, 136–143.
- Ito, T., Tashiro, K. *et al.* (2000) Toward a protein-protein interaction map of the budding yeast: a comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. *Proc. Natl Acad. Sci. USA*, **97**, 1143–1147.
- Jaeger, J.A., Turner, D.H. *et al.* (1990) Predicting optimal and suboptimal secondary structure for RNA. *Math. Enzymol.*, **183**, 281–306.
- Jansen, R. and Gerstein, M. (2000) Analysis of the yeast transcriptome with structural and functional categories: characterizing highly expressed proteins. *Nucleic Acids Res.*, **28**, 1481–1488.
- Jelinsky, S.A. and Samson, L.D. (1999) Global response of *Saccharomyces cerevisiae* to an alkylating agent. *Proc. Natl Acad. Sci. USA*, **96**, 1486–1491.
- Jones, D.T. (1998) Do transmembrane protein superfolds exist? *FEBS Lett.*, **423**, 281–285.
- Jones, D.T. (1999) GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *J. Mol. Biol.*, **287**, 797–815.
- Kidd, D. *et al.* (2001) Profiling serine hydrolase activities in complex proteomes. *Biochemistry*, **40**, 4005–4015.
- Klose, J. (1975) Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik*, **26**, 231–243.
- Krogh, A. *et al.* (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.*, **305**, 567–580.
- Lin, J. and Gerstein, M. (2000) Whole-genome trees based on the occurrence of folds and orthologs: implications for comparing genomes on different levels. *Genome Res.*, **10**, 808–818.
- Lipshutz, R.F. S., Gingeras, T.R. and Lockhart, D.J. (1999) High density synthetic oligonucleotide arrays. *Nature Genet.*, **21**, 20–24.
- Lopez, M.F. (2000) Better approaches to finding the needle in a haystack: optimizing proteome analysis through automation. *Electrophoresis*, **21**, 1082–1093.
- MacBeath, G. and Schreiber, S.L. (2000) Printing proteins as microarrays for high-throughput function determination. *Science*, **289**, 1760–1763.
- Mation, D.P., Constabel, P. *et al.* (1990) Alcohol dehydrogenase gene expression in potato following elicitor and stress treatment. *Plant Mol. Biol.*, **14**, 775–783.
- McCarthy, J.E. (1998) Posttranscriptional control of gene expression in yeast. *Microbiol. Mol. Biol. Rev.*, **62**, 1492–1553.
- Mewes, H.W., Frishman, D. *et al.* (2000) MIPS: a database for genomes and protein sequences. *Nucleic Acids Res.*, **28**, 27–40.
- Millar, A.A., Olive, M.R. *et al.* (1994) The expression and anaerobic induction of alcohol dehydrogenase in cotton. *Biochem. Genet.*, **32**, 279–300.
- Molloy, M.P. (2000) Two-dimensional electrophoresis of membrane proteins using immobilized pI gradients. *Anal. Biochem.*, **280**, 1–10.
- Nauchitel, V.V. and Somorjai, R.L. (1994) Spatial and free energy distribution patterns of amino acid residues in water soluble proteins. *Biophys. Chem.*, **51**, 327–336.
- Nelson, R.W., Nedelkov, D. *et al.* (2000) Biosensor chip mass spectrometry: a chip-based proteomics approach. *Electrophoresis*, **21**, 1155–1163.
- O'Farrell, P.H. (1975) High resolution two-dimensional elec-

- trophoresis of proteins. *J. Biol. Chem.*, 250, 4007-4021.
- Pandey, A. and Mann, M. (2000) Proteomics to study genes and genomes. *Nature*, 405, 837-846.
- Qi, S.Y., Moir, A. et al. (1996) Proteome of *Salmonella typhimurium* SL1344: identification of novel abundant cell envelope proteins and assignment to a two-dimensional reference map. *J. Bacteriol.*, 178, 5032-5038.
- Ross-Macdonald, P., Coelho, P.S. et al. (1999) Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature*, 402, 413-418.
- Rost, B., Casadio, R. et al. (1995) Transmembrane helices predicted at 95% accuracy. *Protein Sci.*, 4, 521-533.
- Roth, F.P., Hughes, J.D. et al. (1998) Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. *Nature Biotechnol.*, 16, 939-945.
- Rubin, G.M., Yandell, M.D. et al. (2000) Comparative genomics of the eukaryotes. *Science*, 287, 2204-2215.
- Sali, A. (1999) Functional links between proteins. *Nature*, 402, 25-26.
- Santoni, V., Molloy, M. et al. (2000) Membrane proteins and proteomics: an amour impossible? *Electrophoresis*, 21, 1054-1070.
- Schena, M., Shalon, D. et al. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, 270, 467-470.
- Searls, D.B. (2000) Using bioinformatics in gene and drug discovery. *Drug Discov. Today*, 5, 135-143.
- Senes, A., Gerstein, M. et al. (2000) Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with beta-branched residues at neighboring positions. *J. Mol. Biol.*, 296, 921-936.
- Shapiro, L. and Harris, T. (2000) Finding function through structural genomics. *Curr. Opin. Biotechnol.*, 11, 31-35.
- Sharp, P.M. and Li, W.H. (1987) The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.*, 15, 1281-1295.
- Sherlock, G. (2000) Analysis of large-scale gene expression data. *Curr. Opin. Immunol.*, 12, 201-205.
- Shevchenko, A., Jensen, O.N. et al. (1996) Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc. Natl Acad. Sci. USA*, 93, 14440-14445.
- Smith, K.D. (2000) Probing proteomes-seeing the whole picture? *Nature Biotechnol.*, 18, 1041-1042.
- Tatusov, R.L., Koonin, E.V. et al. (1997) A genomic perspective on protein families. *Science*, 278, 631-637.
- Tekaia, F., Lazcano, A. et al. (1999) The genomic tree as revealed from whole proteome comparisons. *Genome Res.*, 9, 550-557.
- Velculescu, V.E., Zhang, L. et al. (1997) Characterization of the yeast transcriptome. *Cell*, 88, 243-251.
- Wallin, E. and von Heijne, G. (1998) Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.*, 7, 1029-1038.
- Washburn, M.P., Wolters, D. et al. (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnol.*, 19, 242-247.
- Washburn, M.P. and Yates, J.R. 3rd (2000) Analysis of the microbial proteome. *Curr. Opin. Microbiol.*, 3, 292-297.
- Wittes, J. and Friedman, H.P. (1999) Searching for evidence of altered gene expression: a comment on statistical analysis of microarray data [editorial; comment]. *J. Natl. Cancer. Inst.*, 91, 400-401.
- Wolf, Y.J., Brenner, S.E. et al. (1999) Distribution of protein folds in the three superkingdoms of life. *Genome Res.*, 9, 17-26.
- Young, K.H. (1998) Yeast two-hybrid: so many interactions, (in) so little time. *Biol. Reprod.*, 58, 302-311.
- Zhang, M.Q. (1999) Large-scale gene expression data analysis: a new challenge to computational biologists (published erratum appears in *Genome Res.*, 1999, 9, 1156). *Genome Res.*, 9, 681-688.
- Zhu, H., Klemic, J.F. et al. (2000) Analysis of yeast protein kinases using protein chips. *Nature Genet.*, 26, 283-289.
- Zuker, M. (2000) Calculating nucleic acid secondary structure. *Curr. Opin. Struct. Biol.*, 10, 303-310.

200: Clin Cancer Res. 2001 Apr;7(4):999-1009.

FREE full text article at  
clincancerres.aacrjournals.org

Related Articles, Links

**Thymidine kinase, thymidylate synthase, and dihydropyrimidine dehydrogenase profiles of cell lines of the National Cancer Institute's Anticancer Drug Screen.**

Grem JL, Danenberg KD, Behan K, Parr A, Young L, Danenberg PV, Nguyen D, Drake J, Monks A, Allegra CJ.

Developmental Therapeutics Department, Medicine Branch, Division of Clinical Sciences, National Cancer Institute at the National Naval Medical Center, Bethesda, Maryland 20889, USA.

**PURPOSE:** To determine the expression of three targets of 5-fluorouracil (5-FU) and 5-fluoro-2'-deoxyuridine (FdUrd) in human tumor cell lines and to compare these with the 50% growth inhibition concentrations (GI(50)) from the National Cancer Institute database. **EXPERIMENTAL DESIGN:** Thymidine kinase (TK) activity was assessed by conversion of [(3)H]thymidine to [(3)H]TMP. Thymidylate synthase (TS) protein expression was determined by Western analysis. TS and dihydropyrimidine dehydrogenase (DPD) mRNA expression were measured by quantitative reverse transcription-PCR. **RESULTS:** The median (range) for the targets were as follows: 5-FU GI(50), 20.8 microM (0.8-536); FdUrd GI(50), 0.75 microM (0.25-237); TK, 0.93 nmol/min/mg (0.16-5.7); in arbitrary units: TS protein, 0.41 (0.05-2.95); TS mRNA, 1.05 (0.12-6.41); and DPD mRNA, 1.09 (0.00-24.4). A moderately strong correlation was noted between 5-FU and FdUrd GI(50)s ( $r = 0.60$ ), whereas a weak-moderate correlation was seen between TS mRNA and protein expression ( $r = 0.45$ ). Neither TS expression nor TK activity correlated with 5-FU or FdUrd GI(50)s, whereas lines with lower DPD expression tended to be more sensitive to 5-FU. Cell lines with faster doubling times and wild-type p53 were significantly more sensitive to 5-FU and FDURD. **CONCLUSIONS:** The lack of correlation may in part be attributable to the influence of downstream factors such as p53, the observation that the more sensitive cell lines with faster doubling times also had higher TS levels, and the standard procedure of the screen that uses a relatively short (48-h) drug exposure.

PMID: 11309351 [PubMed - indexed for MEDLINE]

67: Regul Pept. 2004 Feb 15;117(2):127-39.

ELSEVIER  
ELSEVIER

Related Articles, Links

### Galanin in pituitary adenomas.

Grenback E, Bjellerup P, Wallerman E, Lundblad L, Anggard A, Ericson K, Aman K, Landry M, Schmidt WE, Hokfelt T, Hulting AL.

Department of Molecular Medicine, Endocrine and Diabetes Unit, Karolinska Hospital, S-17176 Stockholm, Sweden. Eva.Grenback@ks.se

Tumor galanin content was measured in extracts from human pituitary adenomas using a specific RIA method for monitoring human galanin. Twenty-two out of twenty-four tumors contained galanin with notably high levels in corticotroph adenomas, varying levels in clinically inactive tumors, and low levels in GH secreting adenomas. Tumor galanin and ACTH contents were closely correlated in all tumors. In four young patients with microadenomas and highly active Mb Cushing tumor galanin was inversely related to tumor volume. The molecular form of tumor galanin, studied with reverse-phase HPLC, was homogeneous with the majority of tumor galanin coeluting with standard human galanin. In the tumors analysed with in situ hybridization there was a good correlation between galanin peptide levels and galanin mRNA expression. In some tumors galanin mRNA and POMC levels coexisted, in others they were essentially in different cell populations. Levels of plasma galanin-LI were not related to tumor galanin concentration, and galanin levels were in the same range in sinus petrosus close to the pituitary venous drainage as in peripheral blood. Corticotrophin releasing hormone injections in two patients caused ACTH, but no detectable galanin release into sinus petrosus. Our results demonstrate that corticotroph, but not GH adenomas, express high levels of galanin, in addition to ACTH, and that in some tumors both polypeptides are synthesised in the same cell population. However, galanin levels in plasma were not influenced by the tumor galanin content.

PMID: 14700749 [PubMed - indexed for MEDLINE]

**Protein abundancy and mRNA levels of the adipocyte-type fatty acid binding protein correlate in non-invasive and invasive bladder transitional cell carcinomas.**

Gromova I, Gromov P, Wolf H, Celis JE.

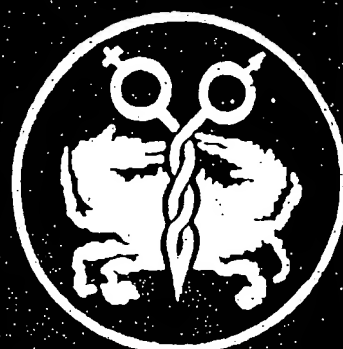
Department of Medical Biochemistry and Danish Centre for Human Genome Research,  
The University of Aarhus, Aarhus C, Denmark.

The adipocyte type fatty acid-binding protein (A-FABP) is a small molecular weight fatty acid-binding protein whose expression correlates both with the grade of atypia and the stage of bladder transitional cell carcinomas (TCCs). To determine if the protein abundancy correlates with the mRNA levels in non-invasive and invasive lesions, we have analysed fresh TCCs (grade II, Ta; grade III, T2-4) by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) and measured the mRNA levels using the reverse transcription linked polymerase chain reaction (RT-PCR). Overall, the results showed a good correlation between protein abundancy and mRNA levels, indicating that the lack of expression of the protein observed in some lesions reflects low levels of transcription of the A-FABP gene rather than translational regulation. In addition, our studies showed that the loss of A-FABP protein observed in some tumors is not compensated by an increase in the skin fatty acid-binding protein PA-FABP, as is the case in the A-FABP knockout mice.

PMID: 9664136 [PubMed - indexed for MEDLINE]

1019-6439  
Received on: 27-23-98  
International Journal of  
Oncology.  
1998

# International Journal of Oncology



ISSN 1019-6439

An international journal devoted to Oncology Research and Cancer Treatment

VOLUME 13, NUMBER 2, AUGUST 1998



# Protein abundance and mRNA levels of the adipocyte-type fatty acid binding protein correlate in non-invasive and invasive bladder transitional cell carcinomas

IRINA GROMOVA<sup>1</sup>, PAVEL GROMOV<sup>1</sup>, HANS WOLF<sup>2</sup> and JULIO E. CELIS<sup>1</sup>

<sup>1</sup>Department of Medical Biochemistry and Danish Centre for Human Genome Research, The University of Aarhus, DK-8000, Aarhus C; <sup>2</sup>Department of Urology, Skejby Hospital, Aarhus, Denmark

Received April 30, 1998; Accepted June 3, 1998

**Abstract.** The adipocyte type fatty acid-binding protein (A-FABP) is a small molecular weight fatty acid-binding protein whose expression correlates both with the grade of atypia and the stage of bladder transitional cell carcinomas (TCCs). To determine if the protein abundance correlates with the mRNA levels in non-invasive and invasive lesions, we have analysed fresh TCCs (grade II, Ta; grade III, T<sub>2-4</sub>) by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) and measured the mRNA levels using the reverse transcription linked polymerase chain reaction (RT-PCR). Overall, the results showed a good correlation between protein abundance and mRNA levels, indicating that the lack of expression of the protein observed in some lesions reflects low levels of transcription of the A-FABP gene rather than translational regulation. In addition, our studies showed that the loss of A-FABP protein observed in some tumors is not compensated by an increase in the skin fatty acid-binding protein PA-FABP, as is the case in the A-FABP knockout mice.

## Introduction

Bladder cancer accounts for 4.7% of all human cancers diagnosed. The spectrum of bladder tumors is broad with various histological types that include transitional cell carcinomas (TCCs), squamous cell carcinomas (SCCs), and adenocarcinomas (1-3). TCCs are by far the most prevalent tumors as they represent nearly 95% of all bladder cancers in

the Western Hemisphere. At first presentation, about 70% of the urinary bladder TCCs are diagnosed as differentiated superficial lesions that are confined either to the mucosa (Ta), or to the underlying connective tissue (T<sub>1</sub>). The rest correspond to highly invasive, poorly differentiated tumors.

Non-invasive TCCs occur as two distinct growth patterns, papillary and non-papillary (flat) lesions (1,2), that display significant differences in their malignant potential and that are believed to originate from different genetic alterations (4-6). Papillary carcinomas usually correspond to low-grade lesions which frequently recur multiple times. These tumors begin as areas of hyperplasia that later undergo a process of dedifferentiation (grades I-IV). Invasive tumors may arise from these lesions, but often are derived from non-papillary carcinoma *in situ* that usually is of high grade at presentation and tend to invade and progress to muscle invasion and metastatic disease.

To date, many attempts have been made to pinpoint genetic changes that underly progression of bladder cancer as well as to identify molecular markers that correlate with tumor progression. Cytogenetic studies and molecular genetic data have shown that chromosomes 3p, 4p, 4q, 5q, 8p, 9p, 9q, 11p, 13q, 14q, 17p and 18q are frequently altered in bladder urothelial tumors (4,5 and refs. therein), and as a whole they have supported the notion that bladder cancer is a multistep process. Recently, Spruck *et al* (6) showed that chromosome 9 alterations occur early during development, while p53 mutations appear later in the process and confer invasive properties. The situation however is reverse in the case of Cis, as a large fraction of these lesions contain p53 mutations (5,6,8,9). Besides pointing towards two divergent pathways of bladder tumor progression, these studies implied that the order in which the genetic changes occur is important in determining the outcome of the lesion.

Assessment of bladder cancer is based on a thorough pathological examination of biopsy material which establishes the histological type of the tumor, its degree of differentiation (grade), and depth of invasion of the bladder wall (staging) (10-12). In spite of strict criteria for the pathological assessment of these lesions, there exist a significant inter-pathologist variation, a fact that emphasises the need for objective markers that may aid their classification. With this in mind, we are exploring the possibility of using proteome (13)

**Correspondence to:** Dr Irina Gromova, Department of Medical Biochemistry and Danish Centre for Human Genome Research, The University of Aarhus, DK-8000, Aarhus C, Denmark

**Abbreviations:** A-FABP, adipocyte type fatty acid-binding protein; PA-FABP, psoriasis associated fatty acid-binding protein; RT-PCR, reverse transcription linked polymerase chain reaction; 2D PAGE, two-dimensional polyacrylamide gel electrophoresis

**Key words:** progression, proteome, protein profiling, A-FABP protein and mRNA levels



expression profiles of these lesions as fingerprints to define their grade of atypia and eventually their stage (3,14). So far, more than 400 tumors of various grades and stages have been analysed by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), and preliminary experiments have shown that even though the overall protein expression profiles of tumors of the same grade and stage are very similar, there are important differences suggesting that morphologically 'identical' TCCs may be further subdivided (1). Of the biomarkers of TCC progression identified so far, the adipocyte-type fatty acid binding protein (A-FABP) is perhaps one of the most interesting as the levels of this polypeptide have been shown to correlate both with the grade of atypia as well as with the stage of the disease (3). Given the putative importance of A-FABP as a progression marker, and since Anderson and Seilheimer (15) recently showed that post-transcriptional regulation of gene expression is a frequent phenomena in higher organisms, we have compared the levels of A-FABP mRNA and protein in non-invasive and invasive bladder TCCs expressing and lacking this protein.

#### Materials and methods

**Tumors.** Fresh bladder tumors were obtained immediately after transurethral resection. The grade and clinical stage of the tumors were determined by the pathologist at the Aarhus Municipal hospital. Clean tumors devoid of blood clots were divided into small pieces for 2D-PAGE and DNA, and RNA preparation. The latter were immediately frozen in liquid nitrogen and store at -80°C until use.

**[<sup>35</sup>S]-methionine labeling and 2D-PAGE.** In a few cases, small tumor pieces were labeled with [<sup>35</sup>S]-methionine as previously described (3). 2D-PAGE was performed according to published procedures (16; see also <http://biobase.dk/cgi-bin/celis>).

**RT-PCR.** Frozen tumor samples were ground to powder in liquid nitrogen and total RNA was isolated using the acid guanidium isothiocyanate/phenol chloroform extraction procedure (17). The samples were treated with RNase-free DNases I (Pharmacia) to eliminate contaminating genomic DNA using the protocols recommended by the supplier. Poly(A)<sup>+</sup> RNA was prepared using Poly (A)<sup>+</sup> Quick columns according to the manufacturer's instructions (Stratagene). The synthesis of cDNA for RT-PCR reactions were carried out using the Gibco BRL SuperScript Kit. Two µg of total RNAs was mixed with oligo-dT11 primer, PCR buffer, MgCl<sub>2</sub> (25 mM), 0.1 M DDT and 10 mM dNTP. The mixture was incubated at 42°C for 5 min followed by the addition of SuperScript II reverse transcriptase and further incubation at 42°C for 50 min. The reaction was terminated by raising the temperature to 70°C for 15 min, followed by additional incubation at 37°C for 20 min in the presence of RNase H to deplete the RNA. Primers for known and cloned genes were purchase from Pharmacia as follows: A-FABP, Upper (from 186-208 bp) 5'-GATCATCAGTGTGAATGGGGAT-3'/lower (from 374-397 bp) 5'-CATCCTCTCGTTTCTCTTTATG-3'; B-actin upper 5'-GAGGTGGCTCTGACTGTACCAC-3'/lower 5'-CTCATTCAGCTCTCGGAACATCTCG-3'.

Table 1. Expression of A-FABP in non-invasive and invasive bladder TCCs:

TCC	Grade/ stage	Level of A-FABP protein <sup>a</sup>	Level of A-FABP mRNA <sup>b</sup>
154	GrlI/Ta	+	++
166-5	GrlI/Ta	-	+
532-1	GrlI/Ta	++++	++++
533-1	GrlI/Ta	+	+
607-1	GrlI/Ta	-	-
692-1	GrlI/Ta	+++	+++
709-1	GrlI/Ta	-	-
763-1	GrlI/Ta	++	++
581-1	GrlI/Ta	+	+
616-1	GrlI/Ta	+	++
428-5	GrlII/T <sub>2</sub> -T <sub>1</sub>	-	-
570-2	GrlII/T <sub>2</sub> -T <sub>1</sub>	-	-
612-3	GrlII/T <sub>2</sub> -T <sub>1</sub>	-	-
711-1	GrlII/T <sub>2</sub> -T <sub>1</sub>	-	-
712-1	GrlII/T <sub>2</sub> -T <sub>1</sub>	-	-
727-1	GrlII/T <sub>2</sub> -T <sub>1</sub>	-	-

<sup>a</sup>The levels of A-FABP protein were determined based on the visual analysis of Coomassie Brilliant Blue stained gels and represent the average estimate of at least two different runs. Tumors scored as positive differed significantly with respect to the levels of the protein, and therefore are indicated with either four (very high), three (high), two (medium) and one (low) cross (see also Fig. 1);

<sup>b</sup>The mRNA levels were determined based on the intensity of Ethidium Bromide stained cDNA bands separated on agarose gel using the Bio-Rad Gel Doc 1000 system and represent the average estimate of at least three independent experiments. Corresponding mRNA levels are represented by crosses as described above.

PCR reactions were performed in a Progene thermal cycler using the Advantage Klen Tag Polymerase (Clontech). The cycling parameters consisted of 30 sec of denaturation at 94°C, with annealing of 30 sec at 60°C for B-actin or at 64°C for A-FABP. The extension time was for 2 min at 68°C for 29-40 cycles with the final extension of 7 min at 68°C. The PCR products were separated on 1.5% agarose gel electrophoresis followed by ethidium bromide staining and photography under UV light.

#### Results

**A-FABP protein levels in non-invasive and invasive TCCs.** One hundred suspected TCCs removed at the Department of Urology, Skejby Hospital, were analysed by high resolution 2D-PAGE and Coomassie Brilliant Blue staining. Of these, 10 grade II, Ta TCCs (Table 1) were chosen to correlate A-FABP protein and mRNA levels as these lesions yielded acceptable protein profiles both in terms of their purity as

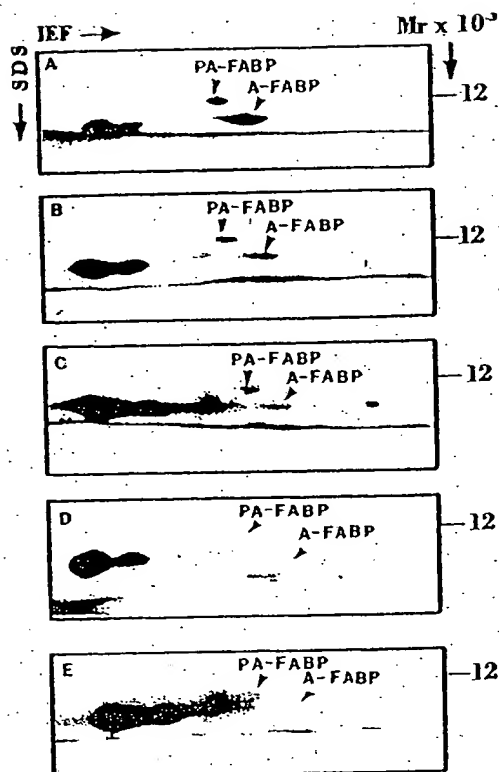


Figure 1. IEF 2D gels of whole cellular extracts from non-invasive and invasive TCCs. A, TCC 532-1; B, TCC 692-1; C, TCC 763-1; D, TCC 709-1 and E, TCC 711-1. Only the relevant area of the gels are shown.

assessed by monitoring for the absence of vimentin (contamination with connective tissue) and desmin (contamination with smooth muscle cells), as well as polypeptide resolution. In addition, reasonable amounts of these tumors were available for mRNA preparation.

Table I shows the levels of A-FABP protein expression in the 10 tumors analysed by 2D PAGE. The data were scored entirely based on the visual analysis of Coomassie Brilliant Blue stained gels and represent an average estimate of at least two different runs. Tumors scored as positive differed significantly with respect to the levels of this protein, and therefore are indicated with either four (very high), three (high), two (medium) and one cross (low). Representative examples of Coomassie stained 2D gels of tumors exhibiting very high (TCC 532-1, Fig. 1A), high (TCC 692-1, Fig. 1B), medium (TCC 763-1, Fig. 1C) and undetectable levels (TCC 709-1 and TCC 711-1 Fig. 1D-E) of A-FABP are shown in Fig. 1 (only the relevant area of the gels are shown).

**A-FABP mRNA levels in non-invasive grade II, Ta TCCs.** Since in many instances only a limited amount of fresh tumor was available, we used RT-PCR to determine the levels of A-FABP mRNA in the ten TCCs analysed by 2D PAGE (Fig. 1). Following amplification, the PCR products were analysed by conventional 1.5% agarose gel electrophoresis and visualised by ethidium bromide staining as shown in Fig. 2. The amount

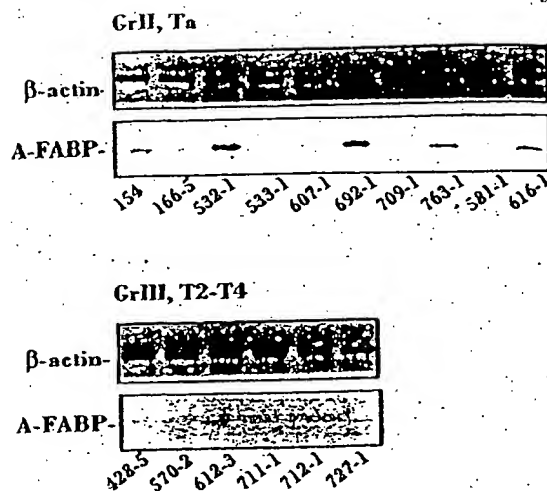


Figure 2. RT-PCR analysis of A-FABP mRNA expression in non-invasive (Gr II, Ta) and invasive TCCs (Gr III, T<sub>2</sub>-T<sub>4</sub>). For RT-PCR analysis, the ss cDNA was synthesized by Reverse Transcriptase using total RNA, and used for RT-PCR amplification. The PCR products were resolved on 1.5% agarose gels and visualised under UV light following ethidium bromide staining. The A-FABP panels show the results of amplifications where the pair of gene specific primers was used to generate the 220 bp DNA fragment. Amplification of A-FABP was obtained after 30 cycles of PCR. The beta-actin panels represent the amplification of the beta-actin gene, which was used as an internal control to confirm that equal amounts of cDNA were used in each reaction.

of cDNA in each lane was normalised using several house-keeping genes so as to achieve a more accurate assessment of the expression of the A-FABP mRNA. As shown in Fig. 2, TCC 532-1 exhibited the highest amount of A-FABP mRNA, followed by TCCs 692-1, 763-1, 616-1, 581-1, 154-1, 166-5 and 533-1. Undetectable levels of A-FABP mRNA were observed in the case of TCCs 607-1 and 709-1 (Fig. 2). Relative mRNA levels for the ten TCCs are given in Table I.

**A-FABP protein and mRNA levels in invasive grade III, T<sub>2</sub>-T<sub>4</sub> TCCs.** Of the invasive TCCs (grade III, T<sub>2</sub>-T<sub>4</sub>) analysed by 2D PAGE only six yielded reasonable protein profiles for further study. As shown in Table I, none of these lesions expressed detectable levels of A-FABP as determined by Coomassie Brilliant Blue staining (Fig. 1E, TCC 711-1). In line with these results, the RT-PCR analysis of these tumors also revealed undetectable level of A-FABP mRNA (Fig. 2, Gr III T<sub>2</sub>-T<sub>4</sub>; Table I).

**Loss of A-FABP protein is not compensated by an increase in PA-FABP.** Recent studies of A-FABP knockout mice have shown that the loss of A-FABP in fat tissue is compensated by an increase in the skin fatty acid-binding protein mall (18). Our studies, however, indicated that the human homologue of mall, PA-FABP (19), did not compensate for the loss of A-FABP either in the non-invasive or the invasive tumors analysed in this study (Fig. 1D and E). In addition, Fig. 3 shows 2D gels of [<sup>35</sup>S]-methionine labeled proteins from two grade II, Ta TCCs (192-4, T<sub>1</sub>; Fig. 3A and 192-4, T<sub>2</sub>; Fig. 3B).

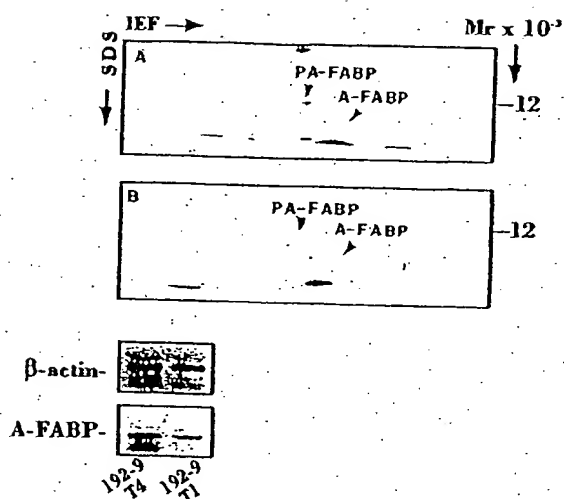


Figure 3. Levels of A-FABP and PA-FABP protein in grade II, T4 tumors resected from the same patient. The two upper panels show the 2D gel autoradiograms of [<sup>35</sup>S]-methionine labeled proteins from TCCs (grade II, T4) resected from the same patient. A, TCC 192-9 tumor 4 and B, TCC 192-9 tumor 1. Only the relevant area of the autoradiograms are shown. The low panel shows the RT-PCR analysis of A-FABP mRNA expression in the same tumors (see also legend to Fig. 2).

which differ significantly in their levels of A-FABP protein and mRNA (Fig. 3, low panel). As shown in Fig. 3, the decrease in A-FABP observed in TCC 129-4, T<sub>1</sub> is not accompanied by an increase in the PA-FABP protein (Fig. 1E).

#### Discussion

Of the TCC progression markers identified to date, A-FABP is perhaps one of the most interesting as its presence correlates both with the grade of atypia ( $p=0.0006$ ) and the stage of the disease ( $p=0.0269$ ) (3). A-FABP is a low molecular weight protein belonging to a cytosolic multigene family of lipid-binding proteins that include heart, liver, intestinal, muscle, brain, skin and epithelial isoforms (20). Members of the FABP family are highly expressed in differentiated cells and show narrow tissue distribution. Their precise function is at present unknown, although there is evidence suggesting that they may play roles in intracellular lipid transport and metabolism, signal transduction (21,22) as well as growth control and differentiation (23). The role in signal transduction has been inferred from the fact that long-chain fatty acids and their metabolites can act as primary and second messengers in specific signalling pathways (24). Recently, it has been shown that A-FABP may play a central role in the pathway that links obesity with insulin resistance, most likely by connecting the fatty acid metabolism with the expression of TNF- $\alpha$  (18). Furthermore, there is evidence indicating that the A-FABP gene contains sequence information necessary for differentiation-dependent expression in adipocytes (25). Our own data derived from the study of TCCs and normal urothelium suggest that A-FABP may be required for normal urothelium differentiation (1), as may be the case for PA-FABP in the skin (19).

Considering the potential prognostic value of A-FABP protein and/or mRNA in TCC progression it was important to determine if the levels of both type of macromolecules correlated both in the non-invasive and the invasive lesions expressing and lacking A-FABP. The need for such correlation was underlined by recent studies of Anderson and Seilhamer (15), who reported a lack of overall correlation between the mRNA and protein levels of 45 rat proteins analysed by 2D PAGE in combination with cDNA arrays. Their data yielded a correlation coefficient of 0.45 which is half way between weak and perfect correlation. Clearly, our data showed a very good correlation between the protein and mRNA levels of A-FABP in all tumors analysed indicating that the loss of A-FABP protein observed in some tumors is not due to post-transcriptional regulation.

Recently, knockout mice carrying a null mutation in the ap2 gene encoding for A-FABP was produced (18). These animals do not show an obvious morphological or metabolic phenotype, but exhibit a 20-fold increase in the levels of the keratinocyte type FABP (*mall*), which may compensate for the loss of the deleted gene (18). The human homologue of the *mall* gene, PA-FABP, was cloned in our laboratory and has been shown to be highly upregulated in psoriatic skin as well as in abnormally differentiated primary keratinocytes (19). PA-FABP is expressed in normal urothelium together with A-FABP (3), and ongoing studies in the laboratory have shown that its level decreases significantly as tumors progress. Interestingly, the studies reported in this article did not reveal a compensatory up- or down-regulation of PA-FABP in the TCCs analysed, supporting the contention that PA-FABP may also play a role in cell growth and differentiation (19).

#### Acknowledgements

We would like to thank Lotte Quist, Gitte Ratz, Jette B. Lauridsen, Bodil Basse, Ariana Celis, Bente Hein and Pamela Celis for expert technical assistance. We also thank Torben Ørntoft for helpful discussion. I. Gromova was supported by a senior fellowship from the Danish Cancer Society. The work was supported by grants from the Danish Cancer Society and the Danish Biotechnology Programme.

#### References

1. Friedell GH, Nagy GK and Cohen SM: The pathology of human bladder cancer and related lesions. In: *The Pathology of Bladder Cancer*. I. Bryan GT and Cohen SM (eds). CRC Press Inc., Boca Raton FL, pp11-42, 1983.
2. Pauli BU, Alroy J and Weinstein RS: The ultrastructure and pathology of urinary bladder cancer. In: *The Pathology of Bladder Cancer*. II. Bryan GT and Cohen SM (eds). CRC Press Inc., Boca Raton FL, pp41-140, 1983.
3. Celis JE, Østergaard M, Basse B, Celis A, Lauridsen JB, Ratz GP, Andersen I, Hein B, Wolf H, Ørntoft TF and Rasmussen HH: Loss of adipocyte-type fatty acid binding protein and other protein biomarkers is associated with progression of human bladder transitional cell carcinomas. *Cancer Res* 56: 4782-4790, 1996.
4. Heney NM, Ahmed S, Flanagan MJ, Frable W, Corder MP, Hafermann MD and Hawkins JR: Superficial bladder cancer: progression and recurrence. *J Urol* 130: 1083-1086, 1983.
5. Simoneau AR and Jones PA: Bladder cancer: the molecular progression to invasive disease (Review). *World J Urol* 12: 89-95, 1994.
6. Spruck CH, Ohnstein PF, Gonzalez-Zulueta M, Esrig D, Miyao N, Tsai YC, Lerner SP, Schutte C, Yang AS and Cote R: Two molecular pathways to transitional cell carcinoma of the bladder. *Cancer Res* 54: 784-788, 1994.

7. Rosin MP, Cairns P, Epstein JJ, Schoenberg MP and Sidransky D: Partial allelotype of carcinoma *in situ* of the human bladder. *Cancer Res* 55: 5213-5216, 1995.
8. Birchmeier W and Birchmeier C: Epithelial-mesenchymal transitions in development and tumour progression. In: *Epithelial-Mesenchymal Interactions in Cancer*. Goldberg ID and Rosen EM (eds). Birkhäuser, Basel, pp1-15, 1995.
9. Sidransky D, von Eschenbach A, Tsai YC, Jones P, Summerhayes I, Marshall F, Paul M, Green P, Hamilton SR, Frost P, *et al*: Identification of p53 gene mutations in bladder cancers and urine samples. *Science* 252: 706-709, 1991.
10. Lopez-Beltran A, Croghan GA, Croghan I, Huben RP, Mentlin C and Gatta JF: Prognostic factors in survival of bladder cancer. *Cancer* 70: 799-807, 1992.
11. Kern WH: The grade and pathologic stage of bladder cancer. *Cancer* 53: 1185-1189, 1984.
12. Jordan AM, Weingarten J and Murphy WM: Transitional cell neoplasms of the urinary bladder. Can biologic potential be predicted from histologic grading? *Cancer* 60: 2766-2774, 1987.
13. Wilkins MR, Sanchez JC, Gooley AA, Appel RD, Humphrey-Smith I, Hochstrasser DF and Williams KL: Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng Rev* 13: 19-50, 1996.
14. Ostergaard M, Rasmussen HH, Nielsen HV, Vorum H, Ørntoft TF, Wolf H and Celis JE: Proteome profiling of bladder squamous cell carcinomas: identification of markers that define their degree of differentiation. *Cancer Res* 57: 4111-4117, 1997.
15. Anderson L and Seilhamer J: A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* 18: 533-537, 1997.
16. Celis JE, Ratz G, Basse B, Lauridsen JB and Celis A: High resolution two-dimensional gel electrophoresis of proteins: isoelectric focusing and non-equilibrium pH gradient electrophoresis (NEPHGE). In: *Cell Biology: A Laboratory Handbook*. III. Celis JE (ed). Academic Press, pp222-230, 1994.
17. Chomczynski P and Sacchi N: Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 162: 156-159, 1987.
18. Hotamisligil GS, Johnson RS, Distel RJ, Ellis R, Papaioannou VE and Spiegelman BM: Uncoupling of obesity from insulin resistance through a targeted mutation in aP2, the adipocyte fatty acid binding protein. *Science* 274: 1377-1379, 1996.
19. Madsen P, Rasmussen HH, Leffers H, Honore B and Celis JE: Molecular cloning and expression of a novel keratinocyte protein psoriasis-associated fatty acid-binding protein (PA-FABP) that is highly up-regulated in psoriatic skin and that shares similarity to fatty acid-binding proteins. *J Invest Dermatol* 99: 299-305, 1992.
20. Veerkamp JH, Paulussen RJ, Peeters RA, Maatman RG, van Moerkerk HT and van Kuppevelt TH: Detection, tissue distribution and (sub)cellular localization of fatty acid-binding protein types. *Mol Cell Biochem* 98: 11-18, 1990.
21. Glatz JF, Vork MM, Cistola DP and van der Vusse GJ: Cytoplasmic fatty acid binding protein: significance for intracellular transport of fatty acids and putative role on signal transduction pathways. *Prostaglandins Leukot Essent Fatty Acids* 48: 33-41, 1993.
22. Spitsberg VL, Matitashvili E and Gorewit RC: Association and coexpression of fatty-acid-binding protein and glycoprotein CD36 in the bovine mammary gland. *Eur J Biochem* 230: 872-878, 1995.
23. Yang Y, Spitzer E, Kenney N, Zschiesche W, Li M, Kromminga A, Muller T, Spener F, Lezius A, Veerkamp JH, *et al*: Members of the fatty acid binding protein family are differentiation factors for the mammary gland. *J Cell Biol* 127: 1097-1109, 1994.
24. Glatz JF, Borchers T, Spener F and van der Vusse GJ: Fatty acids in cell signalling: modulation by lipid binding proteins. *Prostaglandins Leukot Essent Fatty Acids* 52: 121-127, 1995.
25. Hunt CR, Ro JH, Dobson DE, Min HY and Spiegelman BM: Adipocyte P2 gene: developmental expression and homology of 5'-flanking sequences among fat cell-specific genes. *Proc Natl Acad Sci USA* 83: 3786-3790, 1986.

141: Zhonghua Jie He He Hu Xi Za Zhi. 2002 Jun;25(6):337-40.

[Related Articles](#), [Links](#)

**[The pathogenic role of macrophage migration inhibitory factor in acute respiratory distress syndrome]**

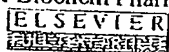
[Article in Chinese]

Guo Y, Xie C.

Department of Respiratory Medicine, First Affiliated Hospital of Zhongshan University, Guangzhou 510080 China.

**OBJECTIVE** To investigate the expression and pathogenic role of macrophage migration inhibitory factor( MIF) in human acute respiratory distress syndrome(ARDS)  
**METHODS** The serum level of MIF in ARDS patients and normal persons were measured by ELISA method. Peripheral blood mononuclear cell (PBMC) MIF expression was determined by flow- cytometry. The expression of MIF mRNA and protein in the lung tissues were detected by using double immuno histochemistry labeling and in situ hybridization. **RESULTS** The serum level of MIF increased significantly in ARDS patients as compared with normal persons ( $P < 0.01$ ). The percentage of PBMC MIF expression was higher in ARDS patients than in normal controls ( $P < 0.01$ ). In situ hybridization and immunohistochemistry showed undetectable or weak MIF mRNA and protein expression in normal lungs. In contrast, there was marked upregulation of MIF mRNA and protein expression in the ARDS lungs. In ARDS macro phages infiltrated the alveolar space and interstitium, most of which also expressed MIF. Infiltrating macrophages were almost restricted to the areas of severe tissue damage. The MIF expression level showed a strong correlation with the number of infiltrating macrophages. **CONCLUSIONS** The serum level of MIF and PBMC MIF expression increased in ARDS patients with enhanced pulmonary MIF expression and macrophage infiltration, which suggests that MIF plays a pivotal role in the pathogenesis of ARDS.

PMID: 12126556 [PubMed - indexed for MEDLINE]



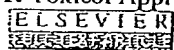
**Restored expression and activity of organic ion transporters rOAT1, rOAT3 and rOCT2 after hyperuricemia in the rat kidney.**

**Habu Y, Yano I, Okuda M, Fukatsu A, Inui K.**

Department of Pharmacy, Kyoto University Hospital, Faculty of Medicine, Kyoto University, Sakyo-ku, Kyoto 606-8507, Japan.

We previously reported that in hyperuricemic rats, renal impairment occurred and organic ion transport activity decreased, accompanied with a specific decrease in the expression of rat organic anion transporters, rOAT1 and rOAT3, and organic cation transporter, rOCT2. In the present study, we investigated the reversibility of the organic ion transport activity and expression of organic ion transporters (slc22a) during recovery from hyperuricemia. Hyperuricemia was induced by the administration of a chow containing uric acid and oxonic acid, an inhibitor of uric acid metabolism. Four days after discontinuance of the chow, the plasma uric acid concentration returned to the normal level, and renal functions such as creatinine clearance and BUN levels were restored, although the recovery of tubulointerstitial injury was varied in sites of the kidney. Basolateral uptake of p-aminohippurate (PAH) and tetraethylammonium (TEA), and both protein and mRNA levels of rOAT1, rOAT3 and rOCT2 in the kidney gradually improved during 14 days of recovery from hyperuricemia. Basolateral PAH transport showed a higher correlation with the protein level of rOAT1 ( $r(2)=0.80$ ) than rOAT3 ( $r(2)=0.34$ ), whereas basolateral TEA transport showed a strong correlation with rOCT2 protein ( $r(2)=0.91$ ). The plasma testosterone concentration, which is a dominant factor in the regulation of rOCT2, was gradually restored during the recovery from hyperuricemia, but the correlation between the plasma testosterone level and rOCT2 protein expression in the kidney was not significant. These results suggest that the regulation of organic ion transporters, rOAT1, rOAT3 and rOCT2, by hyperuricemia is reversible, and the organic ion transport activity restores according to the expression levels of these transporters.

PMID: 15748710 [PubMed - indexed for MEDLINE]



**Regulation of cytochrome P4501A1 in teleosts: sustained induction of CYP1A1 mRNA, protein, and catalytic activity by 2,3,7,8-tetrachlorodibenzofuran in the marine fish *Stenotomus chrysops*.**

**Hahn ME, Stegeman JJ.**

Biology Department, Woods Hole Oceanographic Institution, Massachusetts 02543.

Cytochrome P4501A1 (CYP1A1) is known to play important roles in the activation and detoxification of carcinogens and other toxicants in vertebrate animals, including fish. Although extensively studied in mammalian systems, the regulation of CYP1A forms in other vertebrates is less well understood. We examined the time course and dose-response relationships for induction of CYP1A1 mRNA, protein, and catalytic activity by 2,3,7,8-tetrachlorodibenzofuran (TCDF) in the marine fish *Stenotomus chrysops* (scup). The time course of CYP1A1 induction was determined following a single ip dose (10 nmol/kg) of 2,3,7,8-TCDF. Hepatic ethoxyresorufin O-deethylase activity was increased after 1 day, reached a maximum by 8 days, and was still elevated 14 days after treatment. The content of immunodetectable CYP1A1 protein in liver was elevated on Day 1 and continued to increase through 14 days. CYP1A1 protein content was also strongly induced in heart and gill beginning at 2 days after treatment and extending through Day 14. Hepatic CYP1A1 mRNA was strongly induced by 1 day after dosing and remained elevated through 14 days. The sustained induction of CYP1A1 mRNA by 2,3,7,8-TCDF contrasts with the transient induction seen previously in fish treated with nonhalogenated inducers and most likely reflects differences in persistence of the inducers. Dose-response studies indicated that induction of CYP1A1 mRNA, protein, and catalytic activity occurred following doses of 2,3,7,8-TCDF as low as 0.4 nmol/kg (120 ng/kg), within the range of whole-body contents of this congener measured in fish from contaminated environments. The estimated dose producing half-maximal CYP1A1 induction in scup was approximately 2-10 nmol/kg, suggesting that the sensitivity of these fish to induction may be as great as or greater than that of rats. In contrast to previous results obtained with 3,3',4,4'-tetrachlorobiphenyl (TCB) and beta-naphthoflavone, which appear to inhibit or inactivate CYP1A1 in fish and other vertebrates, there was a good correlation among levels of CYP1A1 mRNA, protein, and catalytic activity in individual fish following various doses of 2,3,7,8-TCDF. The difference in response to 2,3,7,8-TCDF versus 3,3',4,4'-TCB may reflect differences in the inducing potencies of the two compounds relative to their similar potencies as inhibitors of CYP1A1 catalytic activity. In additional studies to evaluate structure-activity relationships for CYP1A1 induction by chlorinated dibenzofurans in fish, scup were treated with 2,3,6,8-tetrachlorodibenzofuran (2,3,6,8-TCDF). At 10 or 50 nmol/kg, 2,3,6,8-TCDF was inactive as an inducer of CYP1A1 mRNA, protein, or catalytic activity. (ABSTRACT TRUNCATED AT 400 WORDS)

PMID: 8048062 [PubMed - indexed for MEDLINE]

**Expression of the pS2 gene in breast tissues assessed by pS2-mRNA analysis and pS2-protein radioimmunoassay.**

Hahnel E, Robbins P, Harvey J, Sterrett G, Hahnel R.

Department of Pathology, University of Western Australia, Queen Elizabeth II Medical Centre, Nedlands.

The expression of the pS2 gene in breast tissues was assessed by measuring pS2-protein using a radioimmunoassay, and by determining pS2-mRNA using Northern blotting. There was a good correlation between the two measurements, indicating that expression of the pS2 gene in breast tissues may be assessed by either method. Since radioimmunoassay is technically easier and more efficient than Northern blotting, radioimmunoassay will be the method of choice in routine applications.

PMID: 1463873 [PubMed - indexed for MEDLINE]



# Breast Cancer Research and Treatment

Marc E. Lippman, MD, editor-in-chief

ITH SOUTHERS LIBRARY  
University of Wisconsin

DEC 17 1992

1306 Linden Drive  
Madison, WI 53706

Kluwer Academic Publishers

# Breast Cancer Research and Treatment

Marc E. Lippman, M.D. <sup>1</sup> (Editor-in-Chief), Gary C. Chamness, Ph.D. <sup>2</sup> / Robert L. Dickson, Ph.D. <sup>1</sup> (Editors),  
C. Kent Osborne, M.D. <sup>2</sup> / Gary M. Clark, Ph.D. <sup>2</sup> (Associate Editors)

<sup>1</sup> Vincent T. Lombardi Cancer Research Center, Georgetown University, Washington DC, USA

<sup>2</sup> University of Texas Health Science Center at San Antonio, San Antonio, TX, USA

## Editorial office address:

Karen S. Cullen, BREA Editorial Office, Kluwer Academic Publishers, 101 Philip Drive, Assinippi Park,  
Norwell, MA 02061, USA; Tel: 617-871-6300; Fax: 617-871-6528; E-mail: Karen@world.std.com.

## EDITORIAL ADVISORY BOARD

George Blumenschein (Arlington,  
Texas)

Gianni Bonadonna (Milan, Italy)

Paul P. Carbone (Madison,  
Wisconsin)

Dean P. Edwards (Denver,  
Colorado)

Evert Engelsman (Amsterdam, The  
Netherlands)

Bernard Fisher (Pittsburgh,  
Pennsylvania)

Edwin Fisher (Pittsburgh,  
Pennsylvania)

Jan-Åke Gustafsson (Stockholm,  
Sweden)

Kathryn Horwitz (Denver,  
Colorado)

Elwood V. Jensen (Hamburg,  
Germany)

V. Craig Jordan (Madison, Wisconsin)

Roger King (London, United Kingdom)

Heinrich Maass (Hamburg, Germany)

Kenneth S. McCarty, Jr. (Durham,  
North Carolina)

Daniel Medina (Houston, Texas)

Henri Rochefort (Montpellier, France)

Richard Santen (Hershey,  
Pennsylvania)

Jeffrey Schlom (Bethesda, Maryland)

Haruo Sugano (Tokyo, Japan)

Jeffrey M. Trent (Tucson, Arizona)

ISSN 0167-6806

All Rights Reserved

© 1992 by Kluwer Academic Publishers

No part of the material protected by this copyright notice may be reproduced or utilised in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without written permission from the copyright owner.

Printed in The Netherlands

*Brief communication*

## Expression of the pS2 gene in breast tissues assessed by pS2-mRNA analysis and pS2-protein radioimmunoassay

Erika Hähnel, Peter Robbins, Jennet Harvey, Gregory Sterrett and Roland Hähnel  
*Department of Pathology, University of Western Australia, Queen Elizabeth II Medical Centre, Nedlands, 6009, Western Australia*

**Key words:** breast tissue, pS2-mRNA, pS2 protein, radioimmunoassay

### Summary

The expression of the pS2 gene in breast tissues was assessed by measuring pS2-protein using a radioimmunoassay, and by determining pS2-mRNA using Northern blotting. There was a good correlation between the two measurements, indicating that expression of the pS2 gene in breast tissues may be assessed by either method. Since radioimmunoassay is technically easier and more efficient than Northern blotting, radioimmunoassay will be the method of choice in routine applications.

### Introduction

Expression of the pS2 gene is controlled by estrogen. This was first described in the MCF-7 breast cancer cell line [1]. pS2 expression has since been reported to be useful as a prognostic indicator [2, 3], although this was not confirmed in another series [4].

pS2 expression may be assessed in tissue homogenates by analysis of pS2-mRNA [5], by radioimmunoassay of the pS2-protein [2], or by immunocytochemical detection of the pS2 protein in tissue sections [5]. It was the aim of this study to establish the correlation between pS2-mRNA and pS2-protein by radioimmunoassay in a series of tissues obtained from mastectomy specimens performed for carcinoma of the breast. Primary breast carcinoma tissue, metastatic carcinoma within axillary nodes, and macroscopically benign breast tissue were examined.

### Materials and methods

#### *Breast tissues*

Tissue specimens from mastectomies performed for carcinoma of the breast were examined. 32 primary breast carcinomas, 10 axillary lymph nodes containing metastatic breast carcinoma, and 20 samples of uninvolved breast tissue were analyzed for pS2 expression.

The primary breast carcinomas were histologically classified using a conventional subclassification. The presence or absence of primary tumour was assessed. The presence of metastatic carcinoma within lymph nodes studied was verified by histological examination of the node remnant after sampling.

'Uninvolved' breast tissue was sampled from sites well removed from the primary breast tumour (usually in another quadrant of the breast), and was selected only if the tissue appeared macroscopically unremarkable. Tissue sampling occurred imme-

*Address for offprints:* R. Hähnel, University Department of Pathology, The Queen Elizabeth II Medical Centre, Nedlands 6009, Western Australia

diately upon arrival of the mastectomy specimen in the laboratory, with minimal delays between removal and sampling.

Tissues for pS2 analysis were snap frozen in liquid nitrogen and stored at  $-70^{\circ}\text{C}$  until processed.

#### *Extraction of RNA and determination of pS2-mRNA*

Details of the procedure have been described in our previous paper [6]. Briefly, the deep-frozen tissue was homogenized in a micro-dismembrator. The homogeneous powder was extracted with guanidiniumisothiocyanatephenolchloroformisoamylalcohol, and RNA was precipitated with isopropanol. The washed RNA pellet was dissolved in SDS and glyoxylated, and the RNA preparation loaded onto agarose gel. After electrophoresis the gel was capillary blotted onto Zeta-probe membranes. Membranes were hybridized overnight with cDNA probes pS2 and 36B4, which were labeled with  $[\alpha^{32}\text{P}]$  dCTP by nick translations. Washed membranes were exposed to Kodak X-omat AR film. Relative intensities of the mRNA bands were assessed visually as not detectable, very weak, weak, medium, strong, and very strong, taking the intensities of the ubiquitous 36B4 bands into account.

#### *Radioimmunoassay of pS2-protein*

Deep frozen specimens were pulverized with a microdismembrator. The tissue powder was suspended in 10 volumes of pH 7.5 phosphate buffer. The homogenate was centrifuged in a refrigerated centrifuge at  $4^{\circ}\text{C}$  for 60 minutes at  $2600\times g$ . The supernatant was removed with a Pasteur pipette, carefully avoiding the fat layer on the top. The protein concentration in the supernatant was estimated by use of the Coomassie dye-binding method [7]. An aliquot of the supernatant was diluted to a protein concentration between 1 and 2 mg/ml before assay of the pS2-protein. In one case the protein concentration of the supernatant was well below 1 mg/ml.

The estimation of the pS2-protein was performed using a solid phase, two-site radioimmunoassay. The kits were bought from CIS Biointernational, Gif-sur-Yvette, France (ELSA-PS2). In this method the molecules of pS2 are sandwiched between two monoclonal antibodies; the first one is coated on the ELSA solid phase, the second one is radiolabeled with 125-iodine. The radioactivity bound to the ELSA is proportional to the concentration of pS2-protein. Details of the procedure are supplied with the kit [8].

#### *Results and discussion*

32 primary breast carcinomas, metastatic breast carcinoma in 10 lymph nodes, and 20 samples of benign breast tissue from mastectomies were investigated. Two of the carcinomas were of the infiltrating lobular type, two were ductal carcinomas *in situ*, one was a multicentric invasive ductal carcinoma, all others were invasive ductal carcinomas.

Examples of pS2 Northern blots have been shown in our previous paper [6] which demonstrate that undegraded pS2-mRNA can be isolated by the method used.

The results of the pS2-protein and pS2-mRNA assays are shown on Fig. 1. There was a good correlation between the two types of results. When pS2-mRNA could not be detected by Northern blot, pS2-protein results were usually below 1 ng/mg protein (22 of 30), or between 1 and 3.7 ng/mg (6 of 30). Two were exceptions (7.7 and 14.6); one of them could have been due to the very low protein content in the cytosol which would lead to a large pS2 value and an associated error. There was no explanation for the other high result. Very weak pS2-mRNA signals on Northern blots corresponded to pS2-protein values between 1.1 and 19.2 with an average of 6.6 ng/mg protein (median 5.7). The mean and median pS2-protein concentration in the tissue with weak pS2-mRNA signals were 14.3 and 10.7 ng/mg protein, respectively. The average pS2-protein concentration increased to 32.7 (median 31.5) ng/mg protein for tissues assessed as medium pS2-mRNA intensity, and to 43.3 (median 53.8) ng/mg protein for tissues with strong or very strong

pS2-mRNA signals. These values should be used as an approximate guide only, since the number of samples in the various groups was fairly small. One-way analysis of variance confirmed that the means of the pS2-protein values in the groups made up according to their pS2-mRNA signal intensity, were significantly different ( $p < 10^{-6}$ ).

If the pS2 gene is expressed, its expression is on average greater in breast carcinomas than in uninvolved breast tissue. If one takes pS2-protein values above 4 ng/mg protein as cut-off, the average pS2-protein in 14 breast cancers was 34.3 (median 35.2), while it was only 18.1 (median 13.8) in 12 uninvolved breast tissue samples. If the cut-off is taken at 10 ng/mg protein, average pS2-protein in breast carcinoma is also about twice the level of uninvolved tissue. There were not enough lymph node metastases which expressed the pS2 gene to allow a comparison with carcinoma or uninvolved breast tissue.

Recent preliminary results of pS2 by radioimmunoassay [9] are similar to ours for breast cancer but considerably lower than our results for normal breast tissue.

The values of the pS2-protein measured obviously depend on the protein used for calibration. We used the pS2-protein standards supplied with the CIS kit, which according to the supplier gave values from 0 to 740 ng/mg protein in a series of 205 breast cancer cytosols. Previously, a different standard had been used for presumably the same series of breast carcinomas [2], and a conversion factor to current standards is given as 2.8 [8].

It was noticed that the correlation between pS2-protein and pS2-mRNA was better in breast carcinoma specimens than in uninvolved breast tissue. This is unexplained, though it could be due to the variable content of cell or tissue types in adjoining parts of a specimen, a variation more likely to occur in our sampling of non-malignant breast tissue compared to sampling of carcinomas. A similar variability in breast carcinoma specimens will probably have a smaller influence on the pS2 results, since the malignant cells – if they do express the pS2 gene – contain more pS2-protein than normal breast.

#### CORRELATION OF pS2-mRNA AND pS2-PROTEIN

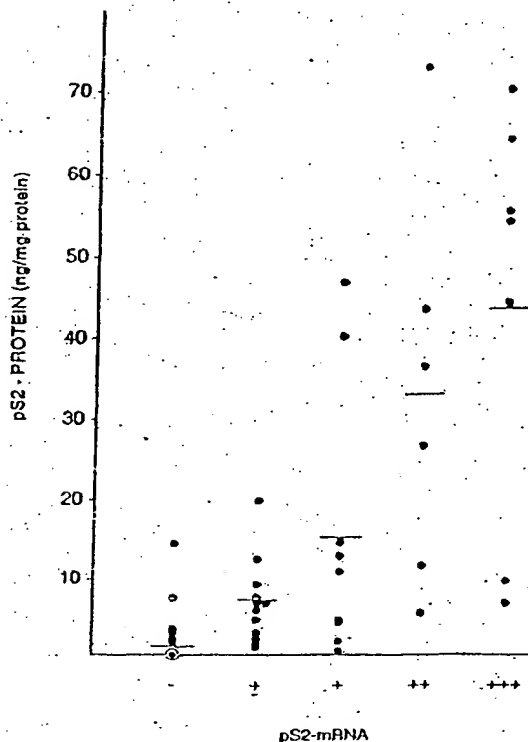


Fig. 1. Correlation between pS2-protein by radioimmunoassay and pS2-mRNA by Northern blot. ⊙ = 22 results below 1. The horizontal lines indicate the mean values.

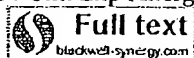
#### Acknowledgements

This investigation was supported in part by a grant from the Sir Charles Gairdner Hospital Research Foundation. The authors wish to thank Professor P. Chambon, Strasbourg, France, for the gift of pS2 and 36B4 cDNAs. A preliminary account of our results was presented at the First Joint Conference of the American Association for Cancer Research and the European Association for Cancer Research, held at Santa Margherita, Italy, 6–9 November, 1991.

#### References

1. Masiukowski P, Breathnach R, Bloch J, Gannon F, Krust A, Cham-

- bon P: Cloning of cDNA sequences of hormone-regulated genes from the MCF-7 human breast cancer cell lines. *Nucleic Acids Res* 10: 7895-7903, 1982
2. Foekens JA, Rio MC, Seguin P, Van Putten WLJ, Fauque J, Nap M, Klijn JGM, Chambon P: Predication of relapse and survival in breast cancer patients by pS2 protein status. *Cancer Res* 50: 3832-3837, 1990
3. Foekens JA, Van Putten WLJ, Portengen H, Rodenburg CJ, Reubi JC, Berns PMJJ, Henzen-Logmans SC, Van Der Burg MEL, Alexieva-Figusch J, Klijn JGM: Prognostic value of pS2 protein and receptors for epidermal growth factor, insulin-like growth factor, and somatostatin in patients with breast and ovarian cancer. *J Steroid Biochem Molec Biol* 37: 815-821, 1990
4. Henry JA, Piggott NH, Mallick UK, Nicholson S, Farndon JR, Westley BR, May FEB: pNR-2/pS2 immunohistochemical staining in breast cancer: Correlation with prognostic factors and endocrine response. *Br J Cancer* 63: 615-622, 1991
5. Rio MC, Bellocq JP, Gairard B, Rasmussen UB, Krust A, Koehl C, Calderoli H, Schiff V, Renaud R, Chambon P: Specific expression of the pS2 gene in subclasses of breast cancer in comparison with expression of the estrogen and progesterone receptors and the oncogene ERBB2. *Proc Natl Acad Sci USA* 84: 9243-9247, 1987
6. Hähnel E, Joyce R, Stern GF, Harvey JM, Hähnel R: Detection of estradiol-induced messenger RNA (pS2) in uninvolved breast tissue from mastectomies for breast cancer. *Breast Cancer Res Treat* 20: 167-176, 1991
7. Bradford M: A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72: 248-254, 1976
8. CIS ELSA-pS2: Immunoradiometric assay of pS2 protein. Package insert, December, 1990
9. Kouyoumdjian JC, Boissier B, Rymer JC, Bagnard G, Rotten D, Levaillant JP, Constancis B, Philippon C, Flouvy C, Thirion B: Determination of several prognostic parameters in human normal breast, benign mastopathies, and adenocarcinomas. *J Tumour Marker Oncol* 6: 111, 1991



**The role of the epidermal growth factor receptor in sustaining neutrophil inflammation in severe asthma.**

Hamilton LM, Torres-Lozano C, Puddicombe SM, Richter A, Kimber I, Dearman RJ, Vrugt B, Aalbers R, Holgate ST, Djukanovic R, Wilson SJ, Davies DE.

Division of Infection, Inflammation & Repair, School of Medicine, University of Southampton, UK.

**BACKGROUND:** The extent of epithelial injury in asthma is reflected by expression of the epidermal growth factor receptor (EGFR), which is increased in proportion to disease severity and is corticosteroid refractory. Although the EGFR is involved in epithelial growth and differentiation, it is unknown whether it also contributes to the inflammatory response in asthma. **OBJECTIVES:** Because severe asthma is characterized by neutrophilic inflammation, we investigated the relationship between EGFR activation and production of IL-8 and macrophage inhibitory protein-1 alpha (MIP-1alpha) using in vitro culture models and examined the association between epithelial expression of IL-8 and EGFR in bronchial biopsies from asthmatic subjects. **METHODS:** H292 or primary bronchial epithelial cells were exposed to EGF or H2O2 to achieve ligand-dependent and ligand-independent EGFR activation; IL-8 mRNA was measured by real-time PCR and IL-8 and MIP-1alpha protein measured by enzyme-linked immunosorbent assay (ELISA). Epithelial IL-8 and EGFR expression in bronchial biopsies from asthmatic subjects was examined by immunohistochemistry and quantified by image analysis. **RESULTS:** Using H292 cells, EGF and H2O2 increased IL-8 gene expression and release and this was completely suppressed by the EGFR-selective tyrosine kinase inhibitor, AG1478, but only partially by dexamethasone. MIP-1alpha release was not stimulated by EGF; whereas H2O2 caused a 1.8-fold increase and this was insensitive to AG1478. EGF also significantly stimulated IL-8 release from asthmatic or normal primary epithelial cell cultures established from bronchial brushings. In bronchial biopsies, epithelial IL-8, MIP-1alpha, EGFR and submucosal neutrophils were all significantly increased in severe compared to mild disease and there was a strong correlation between EGFR and IL-8 expression ( $r = 0.70$ ,  $P < 0.001$ ). **CONCLUSIONS:** These results suggest that in severe asthma, epithelial damage has the potential to contribute to neutrophilic inflammation through enhanced production of IL-8 via EGFR-dependent mechanisms.

PMID: 12580917 [PubMed - indexed for MEDLINE]

349: Arch Biochem Biophys. 1997 Jan 15;337(2):275-83.

ELSEVIER

Related Articles, Links

**Human hepatic microsomal epoxide hydrolase: comparative analysis of polymorphic expression.**

Hassett C, Lin J, Carty CL, Laurenzana EM, Omiecinski CJ.

Department of Environmental Health, University of Washington, Seattle 98105-6099, USA.

Interindividual variation in the expression of human microsomal epoxide hydrolase (mEH) may be an important risk factor for chemically induced toxicities, including cancer and teratogenesis. In this study, phenotypic variability and mEH genetic polymorphisms were examined in a bank of 40 transplant-quality human liver samples. Immunochemically determined protein content, enzymatic activities, polymorphic amino acids, as well as mEH RNA levels were evaluated in parallel. Enzymatic activity was assessed using (+/-)-benzo[a]pyrene-4,5-epoxide at 2 substrate concentrations. The relative hydrolyzing activities obtained using saturating substrate levels were highly correlated ( $r = 0.85$ ) with results derived from limiting substrate concentrations and exhibit approximately an 8-fold range in activity levels across the panel of 40 liver samples. mEH enzyme activity also demonstrated strong correlation ( $r > \text{or} = 0.74$ ) with an 8.4-fold variation determined for mEH protein content within the same samples. However, these protein/activity measurements were poorly correlated ( $r < \text{or} = 0.23$ ) with mEH RNA levels, which exhibited a 49-fold variation. Two common polymorphic amino acid loci in the mEH protein did not exclusively account for variation in enzymatic activity, although this conclusion is confounded by heterozygosity in the samples. These data demonstrate the extent of hepatic mEH functional variability in well-preserved human tissues and suggest that polymorphism of mEH protein expression is regulated in part by posttranscriptional controls, which may include nonstructural regulatory regions of the mEH transcript.

PMID: 9016823 [PubMed - indexed for MEDLINE]



## Review

Early Detection of Lung Cancer: Clinical Perspectives of Recent Advances in Biology and Radiology<sup>1</sup>

Fred R. Hirsch,<sup>2</sup> Wilbur A. Franklin,  
Adi F. Gazdar, and Paul A. Bunn, Jr.

Lung Cancer Program and Departments of Medicine and Pathology,  
University of Colorado Cancer Center, Denver, Colorado 80262  
[F. R. H., W. A. F., P. A. B.]; Department of Pathology, University of  
Texas, Southwestern Medical Center, Dallas, Texas [A. F. G.]; and  
Department of Oncology, Finsen Center, National University  
Hospital, Copenhagen, Denmark [F. R. H.]

## Abstract

Lung cancer is the most common cause of cancer death in developed countries. The prognosis is poor, with less than 15% of patients surviving 5 years after diagnosis. The poor prognosis is attributable to lack of efficient diagnostic methods for early detection and lack of successful treatment for metastatic disease. Most patients (>75%) present with stage III or IV disease and are rarely curable with current therapies. Within the last decade, rapid advances in molecular biology, pathology, bronchology, and radiology have provided a rational basis for improving outcome. These advancements have led to a better documentation of morphological changes in the bronchial epithelium before development of clinical evident invasive carcinomas. This has changed our concept of lung carcinogenesis and emphasized the multistep carcinogenesis approach on several levels. Combined with the technical developments in bronchoscopic techniques, e.g., laser-induced fluorescence endoscope (LIFE) bronchoscopy, we now have improved methods to localize preinvasive and early-invasive bronchial lesions. With the LIFE bronchoscope, a new morphological entity (angiogenic squamous dysplasia) has been recognized, which might be an important biomarker and target for antiangiogenic chemopreventive agents. To reduce the mortality of lung cancer, these new technologies have been taken into the clinic in different scientific settings. The use of low-dose spiral computed tomography in the screening of a high-risk population has demonstrated the possibility of diagnosing small peripheral tumors that are not seen on conventional X-ray. A shift in the therapeutic paradigm from targeting advanced clinically

manifest lung cancer toward asymptomatic preinvasive and early-invasive cancer is occurring. The present article reviews the recent advances in the diagnosis of preinvasive and early-invasive cancer to identify biomarkers for early detection of lung cancer and for chemoprevention studies.

## Introduction

Lung cancer is the most common cause of cancer deaths in the countries of North America and other developed countries, accounting for 29% of all cancer deaths and more deaths than from prostate, breast, and colorectal cancer combined in the United States (1). Lung cancer will be diagnosed in ~170,000 new patients in the United States in the year 2000, and <15% of them will survive 5 years after diagnosis (1). The prognosis for the patients with lung cancer is strongly correlated to the stage of the disease at the time of diagnosis. Whereas patients with clinical stage IA disease have a 5-year survival of about 60%, the clinical stage II-IV disease 5-year survival rate ranges from 40% to less than 5% (2). Over two-thirds of the patients have regional lymph-node involvement or distant disease at the time of presentation (3). The poor prognosis is largely attributable to the lack of effective early detection methods and the inability to cure metastatic disease. The unsatisfactory cure rates supports efforts aimed at early identification and intervention in lung cancer.

Historically, the only diagnostic tests available for the detection of lung cancer in its early stages were chest radiography and sputum cytology. The efficacy of these tests as mass screening tools was evaluated in controlled trials sponsored by the NCI<sup>2</sup> and conducted at Johns Hopkins University, Memorial Sloan-Kettering Cancer Center, and the Mayo Clinic during the 1970s (4-6). The principal goal of these studies was to determine whether a reduction in lung cancer mortality could be achieved by adding sputum cytology testing to annual screening by chest radiography. Results from these trials showed that both tests could detect presymptomatic, early-stage carcinoma, particularly of squamous cell type. Resectability and survival rates were found to be generally higher in the study groups than in the control groups. However, improvements in resectability and survival did not lead to a reduction in overall lung cancer mortality, the most critical end point. A subsequent study of 6346 Czechoslovakian male smokers also found no reduction in lung cancer mortality after dual screening by chest radiography

Received 6/29/00; revised 10/16/00; accepted 10/30/00.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

<sup>1</sup>Supported by National Cancer Institute Grants CA 58187 from the Specialized Program of Research Excellence (SPORE) Lung and CA 85079 from the Lung Cancer Biomarkers and Chemoprevention Consortium.

<sup>2</sup>To whom requests for reprints should be addressed, at University of Colorado Cancer Center, Department of Pathology, University of Colorado Health Sciences Center 4200 East Ninth Avenue, B216, Denver, Colorado 80262. E-mail: Fred.Hirsch@UCHSC.edu

<sup>2</sup>The abbreviations used are: NCI, National Cancer Institute; CIS, carcinoma *in situ*; CT, computed tomography; ASD, angiogenic squamous dysplasia; TSG, tumor suppressor gene; LOH, loss of heterozygosity; hnRNP, heterogeneous nuclear ribonucleoprotein; SCLC, small cell lung carcinoma; BAL, bronchoalveolar lavage; SCLC, small cell lung carcinoma; WLB, white light bronchoscopy; LIFE, laser-induced fluorescence endoscope; ELCAP, Early Lung Cancer Action Project; PET, positron emission tomography; FDG, [<sup>18</sup>F]fluoro-2-deoxyglucose.

and sputum cytology (7). The negative results from these screening studies lead the NCI and other health policy and research groups to conclude that mass screening programs involving periodic sputum cytological evaluation and chest radiographs could not be justified. However, controversies in the methodology and interpretation of the data from these studies have later been extensively discussed (8, 9). One additional study of annual chest X-ray screening is currently being conducted by the NCI; The Prostate-, Lung-, Colorectal-, and Ovarian (PLCO) screening trial. This trial includes individuals 55–74 years old, but they are not selected for this trial on the basis of high risk for lung cancer (e.g., smoking history with >20 pack-years).

The failure of clinical trials to demonstrate the efficacy of sputum cytology and chest radiography as mass screening tools has resulted in a search for better diagnostic approaches for early lung cancer detection that take advantage of recent developments in molecular biology, gene technology, and radiology (10). Furthermore, as has been the case for mammography screening for breast cancer, it has also been important to identify risk groups for lung cancer.

Although, much is known about predisposing factors, natural history, and the outcome based on histology and stage, our understanding remains very incomplete in many areas. What are the early premalignant changes molecularly, biochemically, and morphologically? Which changes are reversible and which are not? What research tools are available to provide answers to these questions? The identification of preinvasive lesions allows for developing promising methods for early intervention (11). The therapeutic paradigm and focus are today shifting from targeting only clinically verified lung cancer as previously toward targeting the premalignant and early-malignant lesions. Furthermore, the prospect of lung cancer screening has today become more meaningful as a consequence of recent developments in biology and radiology and better possibilities to define high-risk populations most suitable for lung cancer screening (12).

The present article will focus on the clinical perspectives of our biological knowledge of premalignant and early-malignant lesions and the potential of the recent technological advancement for early diagnosis of lung cancer.

#### Pathology of Preinvasive and Early Invasive Bronchial Lesions

Most of the efforts to classify lung cancer have been directed toward invasive carcinoma (13). However, better understanding of the pathogenesis of lung cancer aroused renewed interest in morphological abnormalities that fall short of invasive carcinoma but may indicate initiation of carcinogenesis. These morphological abnormalities are referred to as preinvasive lesions and are shown in Fig. 1. The last edition of the WHO classification of lung tumors included the classification of preinvasive lesions as a separate section. Numerous recent studies have indicated that lung cancer is not the result of a sudden transforming event in the bronchial epithelium but a multistep process in which gradually accruing sequential genetic and cellular changes result in the formation of an invasive (i.e., malignant) tumor. Mucosal changes in the large airways that

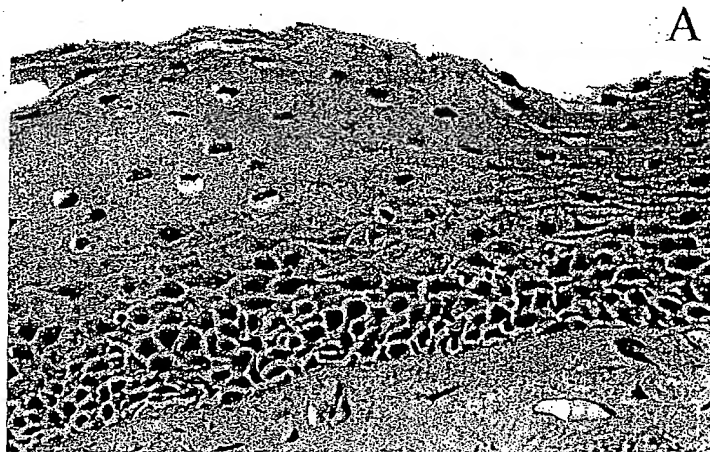
may precede or accompany invasive squamous carcinoma include hyperplasia, metaplasia, dysplasia, and CIS (14). Hyperplasia of the bronchial epithelium and squamous metaplasia have generally been considered reversible, and not premalignant in the sense of squamous dysplasia and CIS (15).

Squamous metaplasia is a common finding, especially as a response to cigarette smoking. Peters *et al.* (16) studied bronchoscopic biopsies from six sites in 106 heavy cigarette smokers; Squamous metaplasia was noted at one or more biopsy sites in approximately two-thirds of the group, and one-fourth showed squamous metaplasia in three or more biopsy sites. The incidence of squamous metaplasia increased with smoking history and was highest in individuals who had smoked more than two packs of cigarettes a day. Auerbach *et al.* (17) noted similar findings in autopsy tissues: basal cell hyperplasia and squamous metaplasia are increased in smokers in proportion to smoking history. Hyperplasia and metaplasia are believed to be reactive changes in the bronchial epithelium, as opposed to true preneoplastic changes (17, 18). The reasons for this include: (a) they are frequently found in association with chronic inflammation, and may be induced by mechanical trauma; (b) they spontaneously regress after smoking cessation; (c) in chronic smokers, the molecular changes present in these lesions are similar to those present in histologically normal epithelium; and (d) there are no reports linking their presence to increased risk for developing lung cancer. In contrast, moderate-to-severe dysplasia and CIS lesions seldom regress after smoking cessation (19).

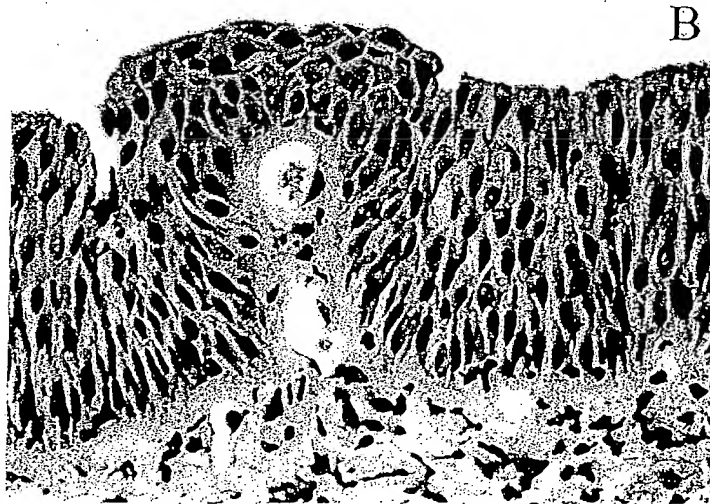
Dysplasia and CIS are changes that frequently precede squamous cell carcinoma of the lung. Saccomanno *et al.* (20) studied more than 50,000 samples from 6,000 men, many of whom had worked in the uranium mining industry. Both smoking and uranium mining (radon exposure) were found to be associated with increased incidence of dysplasia, CIS, and invasive cancer. The studies of Saccomanno *et al.* established that increasing degrees of sputum atypia may be recognized an average of 4–5 years before the development of frank lung carcinoma.

Another question is: which grades of sputum atypia progress to cancer? From the Johns Hopkins cohort of the NCI chest X-ray/sputum screening trial, we know that among individuals with moderate atypia on sputum screening, ~10% developed known cancer up to 9 years later. Among individuals with severe atypia on the sputum screening, >40% developed known cancer during the same time period (21). Although there are data in the literature showing the relationship between sputum atypia and subsequent invasive cancer, there is still very little information about the histological progression in the bronchial mucosa in the high risk populations. In a recent publication, nine patients with CIS were followed with autofluorescence bronchoscopy at regular intervals, and 5 (56%) had progression to invasive cancer despite endobronchial therapy (22). The number of invasive cancers might even have been higher if treatment had not been given. Ongoing studies of high-risk subjects (e.g., the Colorado sputum cohort study) including serial follow-up bronchoscopies will provide evidence related to the frequency of development of invasive lung cancer as it relates to smoking history, airflow obstruction, and sputum atypia.

Since the previous WHO classification was published in



*Fig. 1 A, squamous metaplasia. The cells are widely dispersed, with a regular maturation from the basal region to the top. There is keratinization, and the nuclei/cytoplasmic ratio is low. B, moderate dysplasia with ASD. Hypercellularity of the epithelium with incomplete maturation and micropapillary invasion of capillaries are seen. The nuclei/cytoplasmic ratio is high. C, severe dysplasia. There is marked pleomorphism of the cells with irregularity and prominent nucleoli.*



1981, two nonsquamous lesions have been added to the WHO classification of premalignant lesions: atypical alveolar hyperplasia and diffuse idiopathic neuroendocrine cell hyperplasia (13). Both of these lesions are diagnosed rarely. The former consists of lesions <5 mm in diameter and composed of a peripheral epithelial cell proliferation with minimal cytological atypia or stromal response and resembles bronchioloalveolar carcinoma. The lesion has been seen in lung specimens resected for lung cancer, but no prospective significance of this lesion has been reported. However, this morphological lesion may play a role for the pathogenesis of peripheral lung adenocarcinomas (23, 24). The resolution of spiral CT (currently about 3 mm) approaches the diameter of these lesions, and it is anticipated that atypical alveolar hyperplasia will be increasingly encountered in subjects undergoing this procedure (25). Diffuse idiopathic neuroendocrine cell hyperplasia consists of a patchy increase in the number of well-differentiated neuroendocrine cells in the bronchioles. This process may result in the formation of small carcinoid tumors, and for this reason it is considered "preinvasive." To date, small cell carcinomas have not been associated with this lesion (13).

Recently, the use of fluorescence bronchoscopy (see below) has increased the recognition of dysplastic lesions in the large airways and a new morphological entity, ASD, was identified (26). Dysplasia of bronchial epithelium in "micropapillomatosis" and the possible link between angiogenesis and preinvasive bronchial epithelial dysplasia were recognized as early as 1983 by Muller and Muller (27), who also described the ultrastructure of these lesions. It has been suggested that this angiogenesis, which is recognized as capillary loops projecting into the dysplastic bronchial lining, is responsible for the reduced fluorescence seen in dysplastic lesions by LIFE bronchoscopes (Figs. 1 and 3; Ref. 26). Future prospective studies will show whether this morphological entity is correlated with a progression to lung cancer so as to be a target for the use of antiangiogenic agents for chemoprevention.

In general, there are several questions/problems relating to premalignant lesions, which will be addressed in future studies:

(a) The morphological criteria for premalignant and early-malignant changes, both on sputum cytology and in bronchial biopsies, have to be validated for intra- and interobserver reproducibility.

(b) Uniform and reproducible morphological/cytological criteria have to be published more extensively, and a training set of slides should be available. By the use of Internet technology, this could be more easily facilitated (28).

(c) The correlation of sputum atypia and histological changes in the bronchi in high-risk population is not well defined.

(d) The natural course of preinvasive changes in the bronchi from the high risk subjects needs to be clarified through longitudinal, prospective studies with reference to histological changes in the bronchi. Ongoing longitudinal studies with fluorescence bronchoscopy and multiple biopsies with histology and other biomarkers will define the ability of these markers to assess for risk.

(e) What is the pathology/biology of the small, often peripherally located, tumors (3 mm in diameter), which are more

often diagnosed with newer radiological techniques (e.g., low-dose spiral CT)?

(f) Optimization of the tissue procurement and processing techniques are important. Distinction of reactive from neoplastic processes is usually straightforward, but diagnostic difficulties may arise in the case of (a) inadequate or poorly prepared histological material to evaluate and (b) the presence of cytological atypia in epithelium stimulated by inflammation, viral infection, radiation, or chemotherapy.

(g) DNA array analyses of gene expression: will it be useful? How to collect proper mRNA? Can mRNA extracted from microdissected cells obtained at bronchoscopy be globally amplified and still remain representative of RNA present *in situ*?

### Biology of Lung Carcinogenesis and Potential Early Detection Markers

Lung cancer is the end-stage of multiple-step carcinogenesis, in most cases driven by genetic and epigenetic damage caused by chronic exposure to tobacco carcinogens. The genetic instability in human cancers appears to exist at two levels: at the chromosomal level, including large scale losses and gains; and at the nucleotide level including single or several base changes (29). Lung cancers harbor many numerical chromosomal abnormalities (aneuploidy) and structural cytogenetic abnormalities including deletions and nonreciprocal translocations (30). At least three classes of cellular genes are involved: proto-oncogenes, TSGs, and DNA repair genes. Oncogenic activation often occurs via point mutations, gene amplification, or chromosomal rearrangement, whereas TSGs are classically inactivated by the loss of one parental allele combined with a point or small mutation or aberrant methylation of a target TSG in the remaining allele. Additionally, dysregulated gene expression (either increased or decreased expression) can occur by other, as yet unknown, mechanisms (30). Present studies have not yet confirmed a prominent role for abnormalities of DNA repair genes in lung cancer.

Preneoplastic cells contain several molecular genetic abnormalities identical to some of the abnormalities found in overt lung cancer cells (Fig. 2). These include allele loss at several loci (3p, 9p, 8p, and 17p), *myc* and *ras* up-regulation, cyclin D1 overexpression, p53 mutations, and increased immunoreactivity, *bcl-2* overexpression and DNA aneuploidy (31–35). Allelotyping of precisely microdissected, preneoplastic foci of cells suggests that the earliest changes in the bronchial epithelium is allele loss at chromosome regions 3p, then 9p, 8p, 17p, 5q, and then *ras* mutations (36–39). The biological meaning of LOH is only vaguely understood. Recent evidence suggests that LOH may be a consequence of mitotic recombination, that there is only infrequent physical loss of genetic loci, and that LOH probably precedes chromosomal duplication (40). Allelic loss would thus be significant primarily in the presence of mutation in the retained allele, and gene dosage would not be expected to exert a phenotypic effect in LOH. Some reports have indicated that *ras* activation occurs at early carcinoma stages (34). Histologically normal bronchial epithelium adjacent to cancers has also been shown to have certain genetic losses. Atypical adenomatous hyperplasia, the potential precursor lesion of adenocarcinomas, often have *Ki-ras* mutations (41).

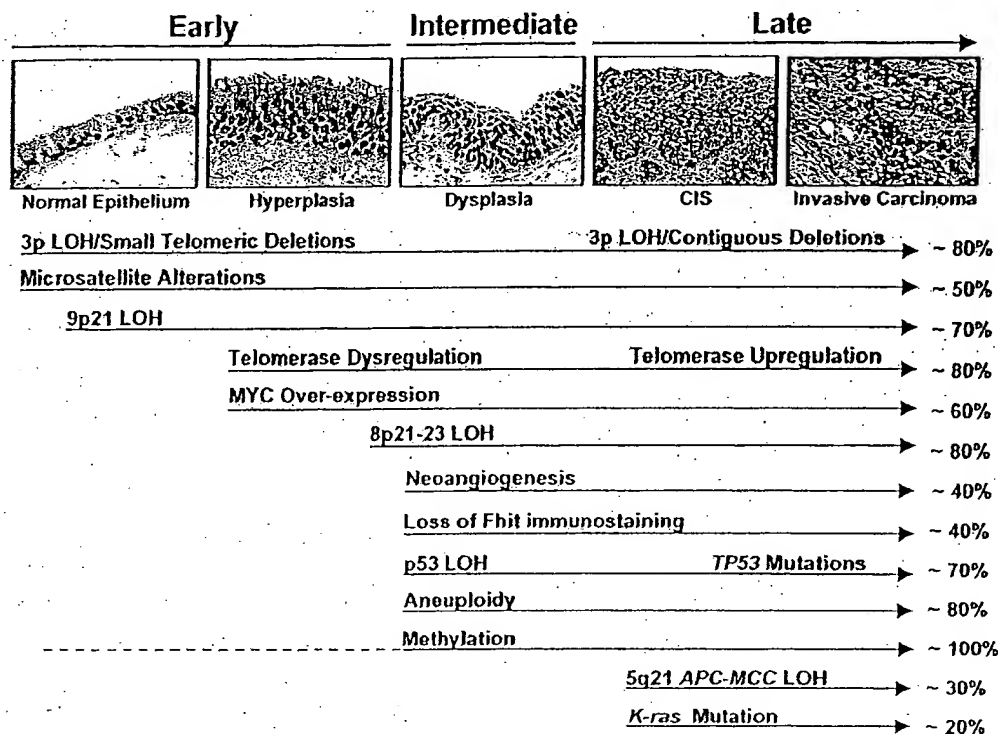


Fig. 2 Sequential changes during lung cancer pathogenesis. Although multiple genetic markers are abnormal in lung cancers, their appearance during the lengthy preneoplastic process varies. The timing of the appearance of these changes has been investigated in bronchial preneoplasia, because sequential sampling of the peripheral lung is technically difficult. Several alterations have been described in histologically normal bronchial epithelium of smokers. Other changes have been detected in slightly abnormal epithelium (hyperplasia, metaplasia) which we do not consider to be true premalignant lesions. These changes are regarded as early changes. Molecular changes detected frequently in dysplasia are regarded as intermediate in timing, whereas those usually detected at the CIS or invasive stages are regarded as late changes. It should be stressed that although there is a usual order, exceptions regarding the timing of onset may occur. Some changes are progressive, such as chromosome 3p deletions. Thus small discrete changes are present early, progressively become more extensive during pathogenesis, and frequently involve all or almost all of the arm in CIS samples. Although allelic loss at the TP53 locus may precede the onset of mutations, data on this sequence are scant. Dysregulation of the RNA component of telomerase (with its appearance in nonbasal cells) is an early event, whereas up-regulation of the gene is a relatively late event.

Molecular changes have been found not only in the lungs of patients with lung cancer, but also in the lungs of current and former smokers without lung cancer (18, 42, 43). These observations are consistent with the multistep model of carcinogenesis and "field cancerization" process, whereby the whole region is repeatedly exposed to carcinogenic damage (tobacco smoke) and is at risk for developing multiple, separate, clonally unrelated foci of neoplasia. The widespread aneuploidy that occurs throughout the respiratory tree of smokers supports this theory (44). However, the presence of the same somatic p53 point mutation at widely dispersed preneoplastic lesions in a smoker without invasive lung cancer indicates that expansion of a single progenitor clone may spread throughout the respiratory tree (45). These molecular alterations might thus be important targets for use in the early detection of lung cancer and for use as surrogate biomarkers in the follow-up of chemoprevention

studies. Detection of these mutant cells should be possible with the different molecular techniques in accessible specimens. The prospects of diagnosing these biological abnormalities in multiple types of clinical specimens are discussed below.

#### Specimens for Clinical Testing: Sputum

Since the 1930s, cytological examination of sputum has been used for the diagnosis of lung cancer (46). Cytological examination of sputa, especially multiple samples, is helpful for the detection of central tumors arising from the larger bronchi (e.g., squamous cell- and small cell carcinomas). Exfoliated cells from peripheral tumors, such as adenocarcinomas, arising from the smaller airways (small bronchi, bronchioles, and alveoli), especially those less than 2 cm in diameter, can be detected only occasionally in sputum samples. This has become of greater importance because the changes in cigarette exposition

(filters and decreased nicotine content) have created an increase in adenocarcinomas and a decrease in squamous carcinomas (47-49). The sensitivity of sputum cytology for early lung cancer is only in the 20%-30% range from screening studies, but by adhering to proper specimen collection, and processing and interpreting criteria, the yield can be substantially improved (50, 51). The data on the reliability of the sputum are conflicting (52-54). Browman *et al.* (52) reported interobserver agreement of 68% for exact and 82% for within-1-category. Holliday *et al.* (54) reported low agreement within observers (27-60%) and across observers (13-50%). Within-1-category intraobserver agreement underwent a two- or 3-fold increase in agreement, which was also the case for interobserver agreement. The variation in intra- and interobserver agreement seems to depend on experience among the cytotechnicians/cytopathologists and the composition of categories studied. A higher degree of agreement is obtained for higher grades of dysplasia (54). Risse *et al.* (55) showed that the ability to detect premalignant conditions is dependent on the number and type of cells present in the deeper airways, suggesting a mode of improvement that is unrelated to observer reliability. MacDougall *et al.* (56) concluded that sputum cytology was too insensitive and insufficiently accurate to be included in the routine work-up of any patient suspected of having lung cancer. To improve the reliability of sputum cytology examinations a simplification of the diagnostic categories from 6 (normal; squamous metaplasia; mild, moderate, and severe atypia; and carcinoma) to 2-3 categories have been proposed (54). Future clinicopathological studies will be required to validate this concept.

To improve the sensitivity of sputum examination as a population-screening tool for the detection of early lung cancer, several approaches are currently under development.

**Immunostaining.** Annual sputum specimens obtained from individuals screened at Johns Hopkins were obtained, and the patients were monitored for 8 years (57). Because the clinical outcome of these patients was known, archival sputum specimens were screened for the presence of biomarkers that could indicate the presence of lung tumors in an early, preinvasive stage. In an attempt to distinguish the pattern of marker expression Tuckman *et al.* (58) studied two monoclonal antibodies. Positive staining predicted subsequent lung cancer approximately 2 years before clinical recognition of the disease, with a sensitivity of 91% and a specificity of 88% (58). One of these antibodies (703 D4) had a higher sensitivity and was later identified as recognizing hnRNP A2/B1 (59). The role of hnRNP A2/B1 overexpression for detecting preclinical lung cancer has been studied in a large high-risk population including 6000 Chinese tin miners who were heavy smokers and who had an extraordinary rate of lung cancer (60). The results from this study indicated that detection of hnRNP A2/B1 overexpression in sputum epithelium cells was 2- to 3-fold more sensitive for detection of lung cancer than standard chest X-ray and sputum cytology methods. The method was particularly effective in identifying early disease (60). The sensitivity was 74% versus 21% for cytology and 42% for chest X-ray. However, the biomarker had a lower specificity (70%) compared with cytology (100%) and chest radiograph (90%). An ongoing clinical trial is evaluating the performance of the A2/B1-protein as a biomarker for the early detection of SPLC. The patients at risk

for SPLC have the highest incidence of lung cancer (2-5%) among asymptomatic populations (61). In this trial, 13 SPLCs were identified by A2/B1, and the sensitivity and specificity were 77-82% and 65-81%, respectively. Among the cases identified as positive by immunocytochemistry and image cytometry, 67% developed SPLC within 1 year (62). Whereas the previous immunocytochemistry studies on material from the older screening material from the NCI-supported screening studies were made on sputum cells cytologically classified with moderately or gravely atypical metaplastic appearance, the latter studies have been done on cytologically "normal appearing" cells. More recently Sueoka *et al.* (63) reported the confirmation of the value of overexpression of hnRNP A2/B1 to detect preclinical lung cancer in Japan. Efforts to improve the sensitivity of hnRNP markers are ongoing (64).

**PCR Techniques.** PCR techniques have been used for the evaluation of molecular biomarkers for early lung cancer detection. In a pilot study with selected patients from the Johns Hopkins Lung Project (JHLP), 8 (53%) of 15 patients with adenocarcinoma or large cell carcinoma were detected by mutations in sputum cells from 1 to 13 months before clinical diagnosis (65). However, the method seemed to be less sensitive than the protein marker described above, and the identification of specific gene abnormalities is further limited by the need to know the specific mutation sequence with which to probe the sputum specimens. Currently, this approach is not practical for screening undiagnosed individuals. Future advances in gene chip technology may permit testing for all possible mutations of common oncogenes and TSGs in clinical specimens of asymptomatic individuals (62).

Microsatellite markers are small repeating DNA sequences found in the noncoding regions of a gene. PCR amplification of these repeat sequences provides a rapid method for assessment of LOH and facilitates the mapping of suppressor genes (66, 67). Microsatellite alterations are extension or deletions of these repeated elements. Detection of microsatellite alterations in histological or cytological specimens may facilitate the detection of clonal preneoplastic or neoplastic cell populations. Although the detection of microsatellite alterations does not indicate the specific genetic change in the tumor, detection of clonal cell populations might serve as a cancer screening marker (65). Identical alterations have been found in lung cancers and corresponding sputum samples demonstrating minimal atypia (68). The *p16* gene is located on the short arm of chromosome 9(9p21) and is frequently mutated or inactivated in tumors and cell lines derived from lung cancer (69, 70). Belinsky *et al.* (71) measured hypermethylation of the CpG islands in the sputum of lung cancer patients and demonstrated a high correlation with early stages of non-small cell lung cancer, which indicated that p16 CpG hypermethylation could be useful in the prediction of future lung cancer. However, prospective studies are needed to evaluate the role of p16 hypermethylation as a marker for early lung cancer detection. Multiple other genes are inactivated by hypermethylation in lung cancer (72), and the detection of hypermethylation may be useful for risk assessment and early diagnosis.

**Computer-assisted Image Analysis.** Computer-assisted image analysis was initially used to detect malignancy-associated changes (e.g., subvisual or nonobvious changes in the

distribution of DNA in the nuclei of histologically normal cells in the vicinity of preinvasive or invasive cancer; 73). In a retrospective analysis of sputum cytology slides, malignancy-associated changes alone correctly identified 74% of the subjects who later developed squamous cell carcinoma (74). The technique has been improved, and recent data showed sensitivities of 75% for stage 0/I lung cancer and 85% for adenocarcinomas with a specificity of 90% (75). This quantitative microscopy technique allows the examining of thousands of cells per slide within a relative short time. Similar techniques have been approved in the United States for cervical cancer screening, and might, in the future, play a role for lung cancer screening. However, no prospective clinical studies have evaluated this technology in a larger lung cancer screening setting.

**High Throughput Technology.** With future advances in gene chip technology, it might become feasible to probe for expression of multiple genes in sputum specimens of asymptomatic individuals. However, this requires a large amount of undegraded RNA from respiratory tract cells. With the high throughput technology, a higher sensitivity might be achieved by using multiple markers at the cost of achieving a lower specificity, which would be undesirable for a screening study.

In conclusion, we need to reevaluate the role of sputum cytology for screening and early detection of lung cancer because of advances in biomarkers and technology. Ongoing studies with standard and biomarker analysis in high-risk groups might change the previous negative attitude and provide a new perspective on sputum cytology as a mass screening tool when applied in a high-risk population. Adding different molecular diagnostic tests gives the possibility for early diagnosis far in advance of clinical presentation. However, validation of the tests in larger prospective studies is necessary, and the individual tests have to be compared with each other to define the role of early diagnosis in the overall management of high-risk subjects. Furthermore, health economic issues have to be considered.

#### Specimens for Clinical Testing: BAL

BAL involves the infusion and reaspiration of a sterile saline solution in distal segments of the lung via a fiberoptic bronchoscope. Ahrendt *et al.* (76) examined a series of 50 resected non-SCLC tumor patients and compared the tumor and BAL with regard to molecular markers including p53 mutations, K-ras mutation, the methylation status of the CpG island of the p16 gene, and microsatellite alteration (Tables 1 and 2). With the possible exception of the test for microsatellite alteration, all of the tests had relatively high sensitivity and could detect mutant cells in the presence of a large excess of normal cells. The frequencies of these changes in the tumors ranged from 27% (for K-ras mutations) to 56% (for p53 mutations). As expected, p53 mutations were more frequent in central (predominantly squamous cell) tumors, and K-ras mutations were more frequent in peripheral (predominantly adenocarcinoma) tumors. The specificity was high (nearly 100%) because, with the exception of microsatellite alterations, the same genetic change in BAL sample as in tumors was always found, but the sensitivity was low, and in only 53% of tumors that contained molecular lesions were the same abnormalities detected in corresponding BAL fluids. Specifically, the tests were least helpful in the

group of patients in whom improved diagnostic abilities are most needed, those with small, peripherally located tumors (77). Unfortunately, the investigators were not able to compare the molecular tests with routine cytopathological analysis of the BAL specimens. The sensitivity of the molecular tests in BAL specimens has to be improved, and we need to know the results from subjects at increased risk (current and former smokers without lung cancer and survivors of previous cancer of the upper respiratory tract) and subjects with chronic lung diseases as well as results from healthy never smokers.

A European group has previously shown that genetic alterations detected in DNA from bronchial lavage of individuals with lung cancer were also found in individuals with no evidence of malignant disease (78), which raises the question about the specificity of such molecular damage in neoplastic conditions. To improve the sensitivity and specificity of detecting allelic imbalance in lung tumors, high throughput PCR-based microsatellite assays have been established (79). In a recent study by Fielding *et al.* (80), the up-regulation of hnRNP A2/B1 was found to be a promising marker in BAL for the detection of premalignant and malignant bronchial lesions with a diagnostic sensitivity of 96% and a specificity of 82%.

It is too early yet to make conclusions as to whether BAL examinations will add to other pathological/molecular biological clinical studies. To obtain diagnostic material for BAL bronchoscopy is required, and we do not have any data that compare BAL examinations with biopsies. Thus, we do not know whether BAL is a valuable adjunct to the biopsies taken under the same bronchoscopy procedure.

#### Specimens for Clinical Testing: Peripheral Blood

For many years scientists have searched for a lung cancer-specific tumor marker that could be detected in peripheral blood. Optimism was raised in the "early" immunocytochemistry era by the use of monoclonal antibodies raised against more-or-less specific epithelial epitopes. In the search for epithelial cells in peripheral blood and bone marrow, monoclonal antibodies against cytokeratin have been used. However, these reactions are clearly not cancer-specific, and some antibodies have been shown to cross-react with normal blood or bone marrow elements (81, 82). Another explanation could be that cells from the macrophage/monocyte system may contain proteins derived from the primary tumor that have undergone necrosis and apoptosis and that these processed proteins are recognized by the antibodies (82). On the basis of "traditional" immunocytochemistry, no markers have been able to detect premalignant or early-malignant disorders based on a peripheral blood sample. However, with the development of DNA technologies, new possibilities have been raised, and, with the use of PCR techniques, some promising reports have been published.

Nanogram quantities of DNA circulating in blood are present in healthy individuals (83, 84). Tumor DNA is also released into the plasma component in increased quantities (85, 86). Thus, the plasma and serum of cancer patients is enriched in DNA, an average four times the amount of free DNA as compared with normal controls (87). In a study by Chen *et al.* (88), a comparison of microsatellite alterations in tumor and plasma DNA was done in SCLC patients, and 93% of the patients with

Table 1 Tissues and other resources for the study of molecular markers

Specimen	Ref.	Comments
Tumor tissue	Numerous	Mixture of cell types, may require microdissection (139). Extensively used for most studies. Alcohol-fixed fine-needle aspirates may be used for mutational and other studies.
Sputum	65, 68, 71, 74	Respiratory cells usually in small minority. Most samples fixed in 'Saccamanno's fixative. Several studies have been performed on these specimens many years later.
Surrogate organ	140	Predominantly squamous epithelial cells. Buccal smears, brushings of tongue or tonsil may be explored as potential surrogate organs resulting from the field effect of tobacco damage of the entire upper aerodigestive tract. This concept needs to be confirmed.
Bronchial brush/wash	141, 142, 143	Predominantly respiratory cells. Fresh, frozen, or alcohol-fixed samples are suitable for multiple studies including FISH.*
Bronchial tissues	42, 43, 45, 144, 145	Usually from bronchial biopsies, but may be obtained from surgical resection specimens. Formalin fixation and paraffin embedding required for histological diagnosis, although EASI preps may permit identification and isolation of subpopulations. Paraffin sections may be used for genotyping polymorphisms, for allelotyping, and for <i>in situ</i> hybridization.
BAL fluids	76, 78, 146, 147, 148	BAL fluids are useful for examining the peripheral airway cells, which are the precursor cells of most adenocarcinomas. Numerous mononuclear cells present. Enrichment of epithelial cells desirable.
Blood components	72, 92, 149	Analysis of circulating tumor cells and genetic material shed by dying tumor cells into the plasma component may yield useful biological and diagnostic information. Gene mutations and presence of epithelial cell markers have been used to detect circulating tumor cells. Gene mutations, allelic loss, microsatellite alterations, and aberrant methylation have been used to identify tumor cell DNA released into the fluid compartment.
Tissue for molecular staging	150, 151	Although little data exists for lung cancers, regional lymph nodes, sentinel lymph nodes, and surgical resection margins have been used in other tumor types for molecular staging.
Tumor cell lines	152, 153	Provide an unlimited self-replicating source of high-quality molecular reagents and have been used for numerous studies. Cell lines may or may not reflect the properties of the tumors from which they were derived (26), although they probably represent cellular subpopulations (27). Aggressive metastatic tumors are more likely to be successfully cultured (28) resulting in skewed data.
Cultures of nonmalignant tissues	154, 155	Epithelial cultures may be useful for studying molecular changes during multistage pathogenesis. Temporary as well as a few immortalized cultures from nonmalignant epithelial cells have been established. B-lymphoblastoid cultures are useful for linkage analysis, for genetic susceptibility studies, and for allelotyping corresponding tumors.
Nonmalignant tissue from patients and from cancer-free relatives	156, 157, 158	Tissues such as buccal smears, tumor-free lymph nodes, and peripheral blood mononuclear cells are useful as controls for linkage analysis, for genetic susceptibility studies, and for allelotyping corresponding tumors.

\* FISH, fluorescence *in situ* hybridization; EASI, epithelial aggregate separation and isolation.

microsatellite alterations in tumor DNA also had modifications in the plasma DNA. However, some patients had LOH only in the tumor DNA. Because most of the microsatellite alterations were similar in tumor DNA and plasma DNA, they concluded that some of the DNA circulating in the blood comes from the tumor. Thus, modifications of circulating DNA can be used as an early detection marker. Detection of aberrant DNA methylation in serum DNA in patients with non-SCLC has been reported (72). Although the number of patients was small and the hypermethylated DNA was found in all stages, it opens up for the possibility to be used as an early lung cancer detection marker. Furthermore, *p53* and *ras* gene mutations have been

detected in the plasma and serum of patients with colorectal cancers (89-91), pancreatic carcinomas (92, 93), and hematological malignancies (94).

In conclusion, the limited direct accessibility of lung carcinomas has led to efforts to identify tumor-associated soluble markers in serum or plasma. Many of the currently recognized soluble markers were first identified as "tumor" markers but, when evaluated in nonneoplastic tissue, have often been found in normal cells as well as in tumors. For early detection of lung cancer, we need more clinical data evaluating these new molecular biological markers from multiple sites, especially in high-risk groups.



Table 2 Molecular approaches for lung cancer investigation

Specimen	Ref.	Comment
Gene mutations	159, 160, 161	Widely used technique, especially for <i>p53</i> and <i>ras</i> genes. Often used to determine the role of a newly discovered gene in the pathogenesis of lung cancer. May be of diagnostic and prognostic significance. Multiple methodologies available.
Allelotyping	18, 158	Useful as a partial substitute for mutational analysis and for determining the chromosomal locations of putative tumor suppressor genes. Widely used to study multistage pathogenesis. Readily performed on formalin-fixed and microdissected tissues. Increasing use of genotyping using automatic sequencers.
Gene expression at RNA and protein level	145, 162, 163, 164, 165, 166	Northern blotting and reverse transcription-PCR are widely used to investigate gene expression. Western blotting often used for detection of protein expression. <i>In situ</i> hybridization for message expression can be performed on paraffin-embedded tissues and, thus, can be used to investigate multistage pathogenesis. Microarray techniques offer promise of examining all or most of the genome but currently require relatively large amount of high-quality RNA from purified cell populations. Sage technique useful for investigation and identification of expressed genes. Similarly, advances in proteomics will permit simultaneous detection of multiple proteins. Numerous immunohistochemical studies of oncogene expression have been used to study multistage pathogenesis. Of particular interest, hnRNP expression on exfoliated epithelial cells in sputum samples may predict for development of cancer.
Molecular cytogenetics	40, 167, 168, 169, 170	<i>In situ</i> hybridization studies of fixed materials or using smears has provided considerable information about numerical and structural changes.
Comparative genomic hybridization	171, 172	Useful for detection of gene amplifications. Less sensitive for the detection of regions of allelic loss.
Morphometric studies	74, 173, 174	May be applied to paraffin-embedded tissues. Useful for determining aneuploidy and for measuring a number of nuclear and cytoplasmic parameters.

#### Specimens for Clinical Testing: Bronchoscopy

WLB is the most commonly used diagnostic tool for obtaining a definite histological diagnosis of lung cancer. Bronchoscopy has major diagnostic limitations for premalignant lesions. Because these lesions are only a few cells thick (0.2–1 mm) and have a surface diameter of only a few millimeters, they rarely are observed as visual abnormalities. Woolner (95) reported that squamous cell CIS was visible to experienced bronchoscopists in only 29% of cases. To address this limitation, fluorescence bronchoscopy was developed. Early studies of fluorescence bronchoscopy entailed the use of fluorescent drugs (hematoporphyrin dyes) that were preferentially retained in malignant tissue (96). Although, studies evaluating this approach did, in fact, show that early invasive and *in situ* cancers could be localized, the detection of dysplasia remained problematic (97–100). Furthermore, the development of photodynamic diagnostic systems was hampered by problems including skin photosensitizing and interference with tissue autofluorescence. To overcome these problems, a new laser photodynamic diagnostic system was developed (101). This system detected tumor-specific drug fluorescence at 630 nm wavelength, which is far from normal tissue autofluorescence (500–580 nm), and interference by autofluorescence from normal tissue should then have been eliminated, but it remained a significant problem (102).

Another approach was developed by Palcic *et al.* (103), who noticed the lack of autofluorescence in the tumor lesions by using blue light (442 nm) rather than white light to illuminate the bronchial surface. They amplified the difference in autofluorescence between normal, premalignant, and tumor tissue for clinical use (103, 104). Using a high-quality-charge coupled device and special algorithm, the LIFE was developed, taking advantage of the principle that dysplastic and malignant tissues reduce autofluorescent signals compared with normal tissue (Fig. 3).

Several studies have been performed comparing the diagnostic specificity and sensitivity of LIFE bronchoscopy versus WLB in diagnosing preinvasive and early-invasive lesions (105–109; Table 3). Most of the studies reported a higher diagnostic sensitivity of LIFE bronchoscopy in the detection of premalignant and early-malignant lesions at the cost of lower specificity (i.e., more false-positive results). In most of these studies, lesions with moderate dysplasia or worse were the target of the study and rated as "positive." The prevalence of preinvasive and early lung cancer varies widely from one study to another, from 20.2% (105) to 65.8% (102). The explanation might be beyond the risk profile of genetic variations or different levels of experience among the endoscopists as well as the pathologists involved. Furthermore, there seems to be a training effect in using the LIFE bronchoscope, which has been demon-

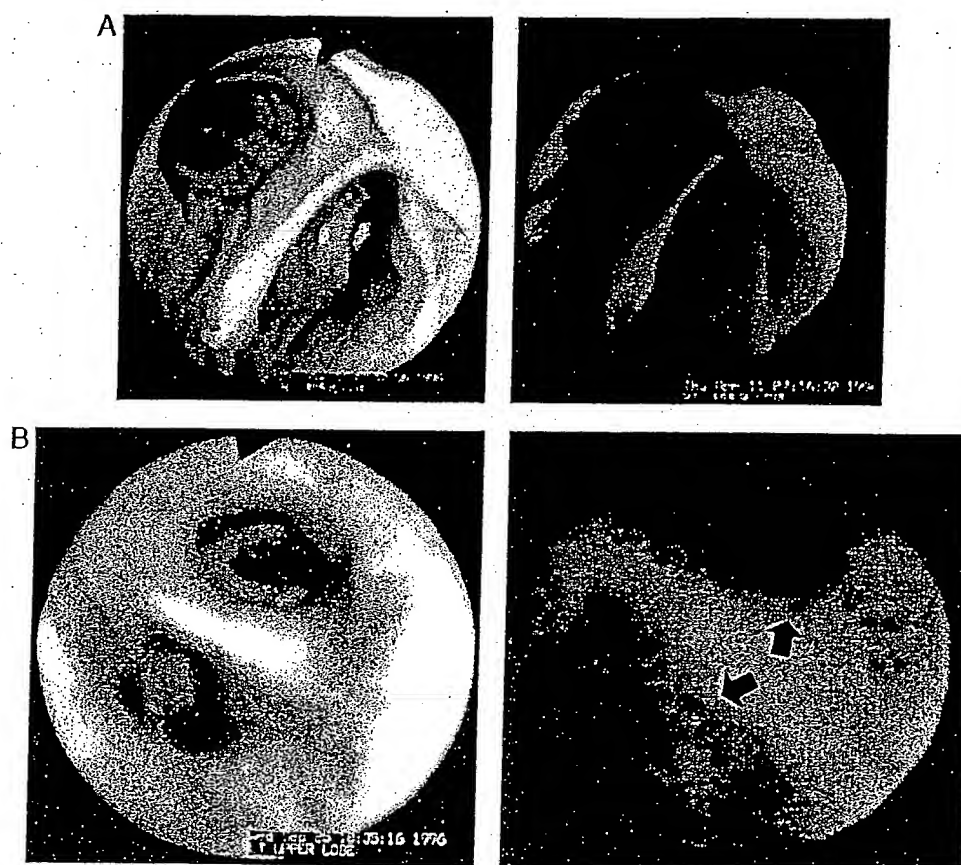


Fig 3. A, normal WLB and normal LIFE bronchoscopy. B, WLB shows inflammatory changes in the bronchial mucosa but no suspicion of malignancy (left). LIFE bronchoscopy shows diffuse reduced autofluorescence (visualized by diffuse red-brownish colorization; arrows). Biopsy demonstrated diffuse severe dysplasia.

strated by Venmans *et al.* (107). In their study, the diagnostic sensitivity increased from 67 to 80% when comparing the first and the second half of the study. The use of the LIFE device in conjunction with WLB improved the detection rate of preneoplastic lesions and CIS significantly (Table 3). Kurie *et al.* (106) looked for more subtle tissue transformation, but their study included few patients with moderate dysplasia or worse. No improvement in the evaluation of metaplasia index was observed by the use of LIFE bronchoscopy. Thus, differences in the study population might explain the different conclusion. There are still no clinical studies with sufficient long-term data showing that moderate dysplasia is the most relevant clinical predictor of eventual malignancy. Limitations in making conclusions from the existing studies are also the potential methodological bias related to the order in which the different bronchoscopy procedures are done and whether the same examiner has performed both procedures. To address these issues, a

prospective randomized study between LIFE bronchoscopy and WLB was done at the University of Colorado Cancer Center. The study design included a randomization with regard to the order of procedure as well as the order of the individual bronchoscopist (109). The order of the procedure and of the individual bronchoscopist did not affect the results. The study also demonstrated a significantly higher sensitivity in detecting premalignant lesions visualized by the LIFE, but at the cost of a lower specificity (109). The reason for the low diagnostic specificity found with the LIFE bronchoscopy in more abnormal foci with the LIFE bronchoscope, with the consequence that a larger number of biopsies were taken and, thus, there was a higher risk of more false-positive results. The use of LIFE bronchoscopy has led to the identification of a new morphological entity, the ASD, which is described above. In a recent morphological study anepidysplastic changes were frequently found in preneoplastic

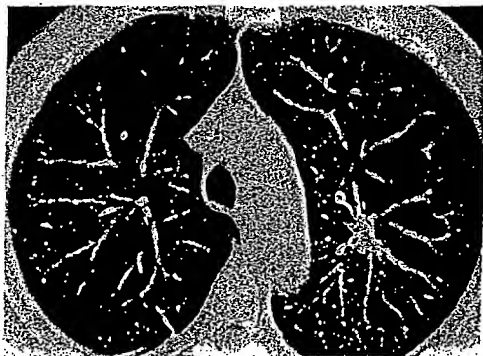


Fig. 4 Seventy-one-year-old man with a spicular nodule in upper left lobe demonstrated on low-dose helical CT (picture), but not visible on chest X-radiography. CT-guided biopsy showed adenocarcinoma.

and early-malignant lesions in the bronchi (26). The morphological entity has been confirmed in preneoplasias among smokers, and the perspectives of this finding have been extensively discussed (110). The prognostic significance of this morphological entity is currently studied in ongoing long-term follow-up studies. Future studies have to evaluate the role of ASD as a biomarker for early lesions and whether it can be used as a marker for treatment effect or therapeutic target for chemoprevention.

The LIFE bronchoscope may play an important role in the screening and follow-up of subjects at high risk of developing lung cancer. At this stage, however, it is unknown whether the LIFE bronchoscope will lead to a reduction in lung cancer mortality. There are also no data on cost-effectiveness and cost-benefit analyses available for this new diagnostic procedure. The use of the LIFE bronchoscope may also in the future be extended to other indications, e.g., patients staged as having resectable lung cancer on one side. Whether LIFE bronchoscopy of the contralateral lung will disclose abnormalities, which would change the therapeutic decision, is not yet reported.

#### Recent Advances in Radiology

The previous NCI-sponsored screening trials failed to demonstrate any reduction in the lung cancer mortality by sputum cytology and yearly chest radiography as mass screening tools for lung cancer screening. Limitations of design and execution of the studies, however, have been discussed extensively (8, 111, 112). An extended follow-up (median, 20.5 years) of the Mayo Lung Project was recently published (113). There was still no difference in lung cancer mortality between the intervention arm and the control arm (4.4 versus 3.9 deaths per 1000 person-years). However, the median survival for patients with resected early-stage disease was 16.0 years in the intervention arm versus 5.0 years in the usual-care arm ( $P < 0.05$ ). The latter findings have raised the question as to whether some small lesions with limited clinical relevance may have been identified in the intervention arm, and the question of "overdiagnosis" was discussed in accompanying editorials (114).

Mass screening for lung cancer has been performed in Japan for many years and has been performed in over 500,000 people in about 80% of the local communities (115). Sobue *et al.* (116) observed that annual clinic-based chest X-ray screening for lung cancer in Japan showed reduced lung cancer mortality by about one-fourth among individuals who underwent screening once a year. In this screening program, the relative odds ratio of dying from lung cancer within 12 months was 0.535 and in the 12–24-month period was 0.638 (117). However, many studies have focused on the pitfalls in the detection of abnormalities by radiography (118–122). The limit of chest radiographic sensitivity for nodule detection is roughly 1 cm in diameter, by which time the tumor has over  $10^9$  cells and may already have violated bronchial epithelium and vascular epithelium. CT has been shown to be more effective in the detection of peripheral lung lesions compared with plain radiography or conventional tomography of the whole lung (123, 124).

Spiral CT scan is a relatively new technology with the ability to continuously acquire data resulting in a shorter scanning time, a lower radiation exposure, and improved diagnostic accuracy compared with those of plain radiography (125–127). Spiral CT allows the whole chest to be imaged in one or two breath-holds, reducing motion artifacts and eliminating respiratory misregistration or missing nodules. Although there is greater radiation exposure with CT than with chest radiography, low-dose techniques (lower mA of 30–50 compared with 200 for conventional CT) have achieved calculated exposure doses that are 17% that of conventional CT and 10 times that of chest radiographs. Further reduction in radiation dose while maintaining diagnostic accuracy is a topic of current research. Furthermore, for the baseline screening, low-dose spiral-CT-scan i.v. contrast is not administered. Nodules as small as 1–5 mm can be shown with modern spiral CT technology (25, 128). The obvious advantages with this new technology led some groups in Japan and in the United States to look to low-dose spiral CT as a tool for screening (Refs. 129–131; Tables 4 and 5).

In a Japanese report, spiral CT scans and chest radiographs were done twice a year in 1369 individuals (129). Peripheral lung cancer was detected in 15 (0.3%) of 3457 examinations, and, among the 15 lung cancer cases detected, the results of chest X-ray were negative in 11 (73%), and the tumors were detected only by low-dose spiral CT. The detection rates of low-dose spiral CT and chest X-ray were 0.43% (15 of 3457 examinations) and 0.12% (4 of 3457 examinations), respectively. Furthermore, 14 (93%) of the 15 lung cancers were stage I disease. The histology showed that 11 of the 15 lung cancer cases were adenocarcinoma, and 4 had squamous cell carcinoma. The effective exposure dose with spiral CT scan in that study was calculated to about one-sixth that of conventional CT.

The ELCAP in New York was designed to determine: (a) the frequency with which nodules were detected; (b) the frequency with which detected nodules represent malignant disease; and (c) the frequency with which malignant nodules are curable (131). In the ELCAP study, 27 lung cancers were found among 1000 subjects screened. Among the 27 patients with cancer, 85% had stage I disease (Table 5).

Another population-based study on low-dose CT screening has been published by Sone *et al.* (130), using a mobile low-dose spiral CT scanner. The detection rate was 0.48% (i.e., 4–5

Table 3 Bronchoscopy versus WLB in diagnosing premalignant and early-malignant lesions

Author	No. of biopsies	Sensitivity				Specificity				Predictive values					
		LIFE+		Relative sensitivity		LIFE+		Relative specificity		PPV <sup>a</sup>	NPV	PPV	NPV	PPV	NPV
		WLB	LIFE	WLB	LIFE+WLB	WLB	LIFE	WLB	LIFE+WLB	WLB	LIFE+	LIFE	LIFE	WLB	WLB
Lam <i>et al.</i> (105)	700	0.67	NR	0.25	6.3 (2.7) <sup>c</sup>	0.66	NR	0.90	NR	0.33	0.89	NR	NR	0.39	0.83
Kurie <i>et al.</i> <sup>b</sup> (106)	234	NR	0.38	NR	NR	NR	0.56	NR	NR	NR	NR	0.16	0.81	NR	NR
Vennans <i>et al.</i> (107)	139	NR	0.89	0.78	1.43	NR	0.61	0.88	NR	0.20	NR	0.14	0.99	0.32	0.98
Vermulen <i>et al.</i> (108)	172	0.93	NR	0.25	3.75	0.21	NR	0.87	NR	0.13	0.96	NR	NR	0.19	0.90
Kennedy <i>et al.</i> (109)	394	0.79	0.72	0.18	4.4	0.3	0.43	0.78	0.38	0.21	0.85	0.25	0.87	0.17	0.80

<sup>a</sup> PPV, positive predictive value; NPV, negative predictive value; NR, not reported.<sup>b</sup> Based on reference pathologist.<sup>c</sup> If invasive carcinoma is included.

Table 4 Results from three population-based screening studies with low-dose spiral CT (LDCT)

Authors	No. of individuals studied	True positive n	False positive <sup>a</sup> %	Predictive value %	Detection rate %			Age incl. yr
					LDCT	X-ray	Pack-yr	
Kaneko <i>et al.</i> (129)	1369	15	15.6	6.6	0.43	0.12	>20	>50
Sone <i>et al.</i> (130)	3967	19	5.0	8.8	0.46-0.5		>30 <sup>b</sup>	40-74
Henschke <i>et al.</i> (131)	1000	27	20.1	11.6	2.7	0.70	>10 <sup>c</sup>	>60

<sup>a</sup> Defined as individuals with "test-positive," in whom further workup gave no suspicion of malignancy.<sup>b</sup> The study also included a group of nonsmokers.<sup>c</sup> Average = 45 (not reported in the other studies).

Table 5 Histology, stage, and size of primary lung cancer detected by low-dose spiral CT

Author	No. of cancers/ No. screened	Histology %			TNM %				Size (mm)				
		Adeno*	Squam.	Other	I	II	III	IV	Average	Range	<10	11-20	>21
Kaneko <i>et al.</i> (129)	15/1369 (1.1%)	73	17		93		7		12	8-18			
Sone <i>et al.</i> (130)	19/5483 (0.3%)	63	5	32	84			16	17	6-47	4	14	3
Henschke <i>et al.</i> (131)	27/1000 (2.7%)	67	3	30	85	4	11				15	8	4

<sup>a</sup> Adeno, adenocarcinoma; Squam., squamous cell carcinoma; TNM, tumor-node-metastasis.

cases per 1000 examinations). Surprisingly, there was no difference in the detection rate among smokers (0.52%) versus nonsmokers (0.46%). The results from the three population-based studies are summarized in Tables 4 and 5. The conclusion from these studies is that 85% of the lung cancers detected by low-dose CT were in stage I, offering improved possibility for curative treatment and better prognosis in general. However, the issue of "false-positive" scans has to be taken into consideration. Thus far, up to 20% of the participants with nodules on the scan had no malignancy during the follow-up period. The possibility that the cancers found represent incidental cancers as in the Mayo Lung Project must also be considered (114). The results from these studies confirm the expectation that low-dose CT increases the detection of small noncalcified nodules and, that lung cancer at an earlier and more curable stage are detected. The mobile CT screening study by Sone *et al.* (130) showed that low-dose CT increased the likelihood of detection of malignant disease 10 times as compared with radiography. The overall rate of malignant disease was lower in the Japanese studies (129, 130) compared with the ELCAP study (Ref. 131; detection rates 0.43-0.48% versus 2.7%). This could be because the Japanese studies screened individuals from the general population ages

40-74, whereas ELCAP screened people at high risk, ages  $\geq 60$ , with a tobacco history of at least 10 pack-years. Thus, as expected, the risk of the population to be screened affects the rate of cancer detection.

Questions remaining to be answered include: (a) what are the diagnostic sensitivity and specificity of this procedure; and (b) does screening reduce lung cancer mortality? The spiral CT has not been as sensitive for small central cancers as it is for small peripheral cancers (129, 131). Minute nodules of lung cancer that are near the threshold of detectability may be overlooked at spiral CT screening (132). A prospective study of the diagnostic sensitivity of spiral CT has recently shown that the diagnostic sensitivity exceeded the sensitivity of conventional CT in previous reports (25). However, there were limitations in the detection of intrapulmonary nodules smaller than 6 mm and of pleural lesions. Compared with surgery (thoracotomy with palpation of deflated lung, resection, and histology), the sensitivity of spiral CT was 60% for intrapulmonary nodules of  $\leq 6$  mm and 95% for nodules of  $\geq 6$  mm and was 100% for neoplastic lesions  $\geq 6$  mm. Furthermore, a marked difference in the sensitivities of two independent observers was found for nodules smaller than 6 mm, whereas agreement was much better for

6-10-mm nodules (25). Given these promising preliminary clinical results, further research is needed to determine the optimal technique for spiral CT screening, which includes collimation, reconstruction interval, pitch, and viewing methods. Decreasing the slice thickness to 3 mm, monitoring the viewing of examinations, and computer-aided diagnosis have been used to improve the diagnostic capability of spiral CT in the detection of pulmonary nodules (133-136).

Future large scale randomized studies have to confirm whether in fact spiral CT screening will lead to a reduction in lung cancer mortality. In a randomized study, the following questions arise: (a) what is the optimal high-risk group to study and what should be the control arm? (b) what should be the end points (goals) of the studies? The ultimate goal is to reduce the lung cancer mortality. However, although this is a long-term goal, intermediate end points from such studies should be evaluated. The change to more curable stages at diagnosis for the lung cancer patients is one such immediate goal; (c) what is the optimal workup and the morbidity of this program? (d) what is the cost of such a screening program? and (e) what is the false-positive rate of the screening findings? Incorporation of smoking cessation programs should be included in the future design of screening studies because it has been shown that screening with low-dose CT in participants who are still smoking provides substantial motivation for smoking cessation (137).

The studies with spiral CT-scan have demonstrated the superior diagnostic ability in the detection of small peripherally located tumors, most of the malignant ones of adenocarcinoma type of histology. The diagnostic sensitivity of spiral CT for more centrally located tumors (mostly squamous cell carcinoma) is significantly lower than for the peripherally located ones. Through these spiral CT studies, we will learn about the biology, pathology, and clinical course of these small tumors, which might be different from what we know about clinically more evident tumors detected routinely in previous studies.

Because lung cancer is so common, the introduction of any new screening technique in this area has to be underpinned by careful definition of the cost implications and must be justified by compelling evidence. The cost-effectiveness of the spiral CT approach should be assessed by evaluating the rate of over-diagnosing nonmalignant, relatively common abnormalities and comparing CT imaging to other diagnostic technologies.

PET with FDG has recently emerged as a practical and useful imaging modality in the preoperative staging of patients with lung cancer. However, whereas CT is most frequently used to provide additional anatomical and morphological information about lesions, the FDG PET imaging provides physiological and metabolic information that characterizes lesions that are indeterminate by CT. FDG PET imaging takes advantage of the increased accumulation of FDG in transformed cells and is sensitive (~95%) for the detection of cancer in patients who have indeterminate lesions on CT (138). The specificity (~85%) of PET imaging is slightly less than its sensitivity because some inflammatory processes avidly accumulate FDG. The high negative predictive value of PET suggests that lesions considered negative on the study are benign, biopsy is not needed, and radiographic follow-up is recommended. Several studies have documented the increased accuracy of PET compared with CT in the evaluation of the hilar and mediastinal lymph node status

in patients with lung cancer (138). However, the PET resolution is sufficient only for nodules  $\geq 6$  mm and will not be helpful in detecting the very small nodules. Compared with low-dose spiral CT, the FDG PET scan is more expensive and time consuming. The role of PET scan in early diagnosis of lung cancer in an asymptomatic high-risk population is not yet evaluated. However, future studies have to include PET evaluation to define its role in a population screening setting.

## Conclusion

Recent advances in molecular biology and pathology have led to a better understanding and documentation of morphological changes in the bronchial epithelium before development of clinical evident lung carcinomas. Combined with technical developments in radiological and bronchoscopic techniques, these procedures offer great promise in diagnosing lung cancer far in advance of clinical presentation. Any of these individual procedures could be incorporated into the routine management of individuals at risk for developing primary or secondary lung cancer, and for several of these methods, clinical studies are under way. Preliminary reported data are very promising for the early detection of lung cancer. Future studies must incorporate the different methods in a multidisciplinary scientific setting to evaluate the role of the individual method in the overall management for individuals at high risk for developing lung cancer. Several of these tests might diagnose the disease at the stage of clonal expansion before invasive carcinoma has developed. A management and intervention strategy appropriate to that stage of disease have to be developed. Preliminary studies of chemoprevention agents are reported, and new agents based on other biological mechanisms are under development and ready for clinical trials. It is now time to plan clinical trials that evaluate both diagnostic and therapeutic approaches to assess their impact on the incidence of clinical lung cancer.

## Acknowledgments

We thank Drs. Stephen Lam, Vancouver, British Columbia, Canada, and Kavita Garg, University of Colorado Health Sciences Center, Denver, Colorado, for a critical review of the manuscript and Drs. Timothy Kennedy and York Miller for submitting illustrations for white-light and LIFE bronchoscopy.

## References

- Greenlee, R. T., Murray, T., Bolden, S., and Wingo, P. A. Cancer Statistics, 2000. *CA Cancer J. Clin.*, 50: 7-30, 2000.
- Mountain, C. T. Revision in the international system for staging of lung cancer. *Chest*, 111: 1710-1717, 1997.
- Idre, D. C. Chemotherapy of lung cancer. *N. Engl. J. Med.*, 327: 1434-1441, 1992.
- Melamed, M. R., Flehinger, B. J., Zaman, M. B., Heelan, R. T., Partridge, W. A., and Mattini, N. Screening for lung cancer: results of the Memorial Sloan-Kettering study in New York. *Chest*, 86: 44-53, 1984.
- Fontana, R. S., Sanderson, D. R., Woolner, L. B., Taylor, W. F., Miller, W. E., and Muhm, J. R. Lung Cancer Screening: The Mayo program. *J. Occup. Med.*, 28: 746-750, 1986.
- Tockman, M. S. Survival and mortality from lung cancer in a screened population: The Johns Hopkins Study. *Chest*, 59: 324S-325S, 1986.

7. Kubik, A., Parkin, D. M., Khat, M., Erban, J., Polak, J., and Adamec, M. Lack of benefit from semi-annual screening for cancer of the lung: follow-up report of a randomized controlled trial on a population of high-risk males in Czechoslovakia. *Int. J. Cancer*, 45: 26-33, 1990.
8. Fontana, R. S., Sanderson, D. R., Woolner, L. B., Taylor, W. F., Miller, W. E., Muhm, J. R., Bernatz, P. E., Payne, W. S., Paisiolo, P. C., and Bergstralh, E. J. Screening of lung cancer. A critique of the Mayo Lung Project. *Cancer (Phila.)*, 67: 1155-1164, 1991.
9. Strauss, G. M., Gleason, R. E., and Sugarbaker, D. J. Screening for lung cancer. Another look: a different view. *Chest*, 111: 754-768, 1997.
10. Hirsch, F. R., Brambilla, E., Gray, N., Gritz, E., Kelloff, G. J., Linnoila, I., Pastorino, U., and Mulshine, J. L. Prevention and early detection of lung cancer—clinical aspects. *Lung Cancer (Limerick)*, 17: 163-174, 1997.
11. Hong, W. K. Chemoprevention of Lung Cancer. *Oncology (Basel)*, 13 (Suppl. 5): 135-141, 1999.
12. Mulshine, J. L., and Henschke, C. I. Prospects for lung-cancer screening. *Lancet*, 355: 592-593, 2000.
13. Travis, W. D., Colby, T. V., Corrin, B., Shimosato, Y., and Brambilla, E. Histological typing of tumours of lung and pleura. In: L. H. Sobin (ed.), *World Health Organization International Classification of Tumours*, Ed. 3. New York: Springer-Verlag, 1999.
14. Franklin, W. A. Pathology of lung cancer. *J. Thorac. Imaging*, 15: 3-12, 2000.
15. Colby, T. V. Precursor lesions to pulmonary neoplasia. In: C. Brambilla and E. Brambilla (eds.), *Lung Tumors. Fundamental Biology and Clinical Management*, pp. 61-87. New York: Marcel Dekker Inc., 1999.
16. Peters, E. J., Morice, R., Benner, S. E., Lippman, S., Lukeman, J., Lee, J. S., Ro, J. L., and Hong, W. K. Squamous metaplasia of the bronchial mucosa and its relationship to smoking. *Chest*, 103: 1429-1432, 1993.
17. Auerbach, O., Gere, B., Forman, J. B., Petrick, T. G., Smolin, H. J., Muchsam, G. E., Kassouny, D. Y., and Stout, A. P. Changes in bronchial epithelium in relation to smoking and cancer of the lung. *N. Engl. J. Med.*, 236: 97-104, 1957.
18. Wistuba, I. I., Behrens, C., Milchgrub, S., Bryant, D., Hung, J., Minna, J. D., and Gazdar, A. F. Sequential molecular abnormalities are involved in the multistage development of squamous cell lung carcinoma. *Oncogene*, 18: 643-650, 1999.
19. Lam, S., LeRiche, J. C., Zheng, Y., Coldman, A., MacAulay, C., Hawk, E., Kelloff, G., and Gazdar, A. F. Sex-related differences in bronchial epithelial changes associated with tobacco smoking. *J. Natl. Cancer Inst.*, 91: 691-696, 1999.
20. Saccomanno, G., Archer, V. E., Auerbach, O., Saunders, R. P., and Brennan, L. M. Development of carcinoma of the lung as reflected in exfoliated cells. *Cancer (Phila.)*, 32: 256-270, 1974.
21. Frost, J. K., Ball, W. C., Jr., Levin, M. L., Tockman, M. S., Erozan, Y. S., Gupta, K., Eggleston, J. C., Pressman, N. J., Donithan, M. P., and Kimball, A. W. Sputum cytopathology: use and potential in monitoring the workplace environment by screening for biological effects of exposure. *J. Occup. Med.*, 28: 692-703, 1986.
22. Venman, B. J. W., van Boxem, T. J. M., Smit, E. F., Postmus, P. E., and Suteldj, T. G. Outcome of bronchial carcinoma *in situ*. *Chest*, 117: 1572-1576, 2000.
23. Slebos, R. J., Baas, I. O., Clement, M. J., Offerhaus, G. J., Askin, F. B., Hruban, R. H., and Westra, W. H. p53 alterations in atypical alveolar hyperplasia of the human lung. *Hum. Pathol.*, 29: 801-808, 1998.
24. Kitamura, H., Kameda, Y., Ito, T., and Hayashi, H. Atypical adenomatous hyperplasia of the lung: Implications for the pathogenesis of peripheral lung adenocarcinoma. *Am. J. Clin. Pathol.*, 111: 610-622, 1999.
25. Diehrich, S., Senn, M., Leitschig, M. G., Winter, F., Scheid, H. H., Roos, N., and Bongartz, G. Helical CT of pulmonary nodules in patients with extrathoracic malignancy: CT-surgical correlation. *Am. J. Roentgenol.*, 172: 353-360, 1999.
26. Keith, R. L., Miller, Y. E., Gemmill, R. M., Drabkin, H. A., Dempsey, E. C., Kennedy, T. C., Prindiville, S., and Franklin, W. A. Angiogenic squamous dysplasia in bronchi of individuals at high risk for lung cancer. *Clin. Cancer Res.*, 6: 1616-1625, 2000.
27. Muller, K. M., and Muller, G. The ultrastructure of preneoplastic changes in bronchial mucosa. *Curr. Top. Pathol.*, 73: 233-263, 1983.
28. Hirsch, F. R., Gazdar, A. F., Gabrielson, E., Lam, S., and Franklin, W. A. Histopathologic evaluation of premalignant and early malignant bronchial lesions: an interactive program based on internet digital images to improve WHO criteria for early diagnosis of lung cancer and for monitoring chemoprevention studies—a SPOR collaborative project. *Lung Cancer (Limerick)*, 29 (Suppl. 2): 209, 2000.
29. Lengauer, C., Kinzler, K. W., and Vogelstein, B. Genetic instabilities in human cancers. *Nature (Lond.)*, 396: 643-649, 1998.
30. Fong, K. M., Sekido, Y., and Minna, J. D. Molecular pathogenesis of lung cancer. *J. Thorac. Cardiovasc. Surg.*, 118: 1136-1152, 1999.
31. Hirano, T., Franzen, B., Kato, H., Ebihara, Y., and Auer, G. Genesis of squamous cell lung carcinoma: sequential changes of proliferation, DNA ploidy, and p53 expression. *Am. J. Pathol.*, 144: 296-302, 1994.
32. Betticher, D. C., Heighway, J., Thatcher, N., and Hasleton, P. S. Abnormal expression of CCND1 and RB1 in resection margin epithelia of lung cancer patients. *Br. J. Cancer*, 75: 1761-1768, 1997.
33. Satoh, Y., Ishikawa, Y., Nakagawa, K., Hirano, T., and Tsuchiya, E. A follow-up study of progression from dysplasia to squamous cell carcinoma with immunohistochemical examination of p53 protein overexpression in the bronchi of ex-chromate workers. *Br. J. Cancer*, 75: 678-683, 1997.
34. Li, Z. H., Zheng, J., Weiss, L. M., and Shibata D. c-k ras, and p53 mutations occur very early in adenocarcinoma of the lung. *Am. J. Pathol.*, 144: 303-309, 1994.
35. Brambilla, E., Gazzeri, S., Lantuejoul, S., Coll, J. L., Moro, D., Negoescu, A., and Brambilla, C. p53 mutant immunophenotype and deregulation of p53 transcription pathway (Bcl2, Bax, and Waf1) in precursor bronchial lesions of lung cancer. *Clin. Cancer Res.*, 4: 1609-1618, 1998.
36. Hung, J., Kishimoto, Y., Sugio, K., Virmani, A., McIntire, D. D., Minna, J. D., and Gazdar, A. F. Allele-specific chromosome 3p deletions occur at an early stage in the pathogenesis of lung carcinoma. *J. Am. Med. Assoc.*, 273: 558-563, 1995.
37. Kishimoto, Y., Sugio, K., Hung, J. Y., Virmani, A. K., McIntire, D. D., Minna, J. D., and Gazdar, A. F. Allele-specific loss in chromosome 9p loci in preneoplastic lesions accompanying non-small cell lung cancers. *J. Natl. Cancer Inst.*, 87: 1224-1229, 1995.
38. Sugio, K., Kishimoto, Y., Virmani, A. K., Hung, J. Y., and Gazdar, A. F. K-ras mutations are a relatively late event in the pathogenesis of lung carcinomas. *Cancer Res.*, 54: 5811-5815, 1994.
39. Wistuba, I. I., Behrens, C., Virmani, A. K., Milchgrub, S., Syed, S., Lam, S., Mackay, B., Minna, J. D., and Gazdar, A. F. Allelic losses at chromosome 8p21-23 are early and frequent events in the pathogenesis of lung cancer. *Cancer Res.*, 59: 1973-1979, 1999.
40. Varela-Garcia, M., Gemmill, R. M., Rabenhorst, S. H., Lott, A., Drabkin, H. A., Archer, P. A., and Franklin, W. A. Chromosomal duplication accompanies allelic loss in non-small cell lung carcinoma. *Cancer Res.*, 58: 4701-4707, 1998.
41. Westra, W. H., Baas, I. O., Hruban, R. H., Askin, F. B., Wilson, K., Offerhaus, G. J., Slebos, R. J. K-ras oncogene activation in atypical alveolar hyperplasias of the human lung. *Cancer Res.*, 56: 2224-2228, 1996.
42. Wistuba, I. I., Lam, S., Behrens, C., Virmani, A. K., Fong, K. M., LeRiche, J., Saret, J. M., Sivastava, S., Minna, J. D., and Gazdar, A. F. Molecular damage in the bronchial epithelium of current and former smokers. *J. Natl. Cancer Inst.*, 89: 1266-1277, 1997.
43. Mao, L., Lee, J. S., Kurie, J. M., Fan, Y. H., Lippman, S. M., Lee, J. J., Ro, J. Y., Braxson, A., Yu, R., Morice, R. C., Krump, B. L., Khuri, F. R., Walsh, G. L., Hittelman, W. N., and Hong, W. K. Clonal genetic

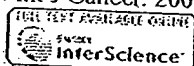
- alterations in the lungs of current and former smokers. *J. Natl. Cancer Inst.*, 89: 857-862, 1997.
44. Smith, A. L., Hung, J., Walker, L., Rogers, T. E., Vuitch, F., Lee, E., and Gazdar, A. F. Extensive areas of aneuploidy are present in the respiratory epithelium of lung cancer patients. *Br. J. Cancer*, 73: 203-209, 1996.
  45. Franklin, W. A., Gazdar, A. F., Haney, J., Wistuba, I. I., La Rosa, F. G., Kennedy, T., Ritchey, D. M., and Miller, Y. E. Widely dispersed p53 mutation in respiratory epithelium: a novel mechanism for field carcinogenesis. *J. Clin. Invest.*, 100: 2133-2137, 1997.
  46. Johnston, W. W., and Elson, C. E. Respiratory tract. In: M. Bibbo (ed.), *Comprehensive Cytopathology*, pp. 325-401. Philadelphia: Saunders, 1997.
  47. Travis, W. D., Travis, L. B., and Devesa, S. S. Lung cancer. *Cancer (Phila.)*, 75 (Suppl. 1): 191-202, 1995.
  48. Wynder, E. L., and Muscat, J. E. The changing epidemiology of smoking and lung cancer histology. *Environ. Health Perspect.*, 103 (Suppl. 8): 143-148, 1995.
  49. Thun, M. J., Lally, C. A., Flannery, J. T., Calle, E. E., Flanders, W. D., and Heath, C. W., Jr. Cigarette smoking and changes in the histopathology of lung cancer. *J. Natl. Cancer Inst.*, 89: 1580-1586, 1997.
  50. Lam, S., and Shibiya, H. Early diagnosis of lung cancer. *Clin. Chest Med.*, 20: 53-61, 1999.
  51. Kennedy, T. C., Proudfoot, S., Piantadosi, S., Wu, L., Saccomanno, G., Petty, T. L., and Tockman, M. S. Efficacy of two sputum collection techniques in patients with air flow obstruction. *Acta Cytol.*, 43: 630-636, 1999.
  52. Browman, G. P., Arnold, A., Levine, M. N., and D'Souza, T. Use of screening phase data to evaluate observer variation of sputum cytodiagnosis as an outcome measure in a chemoprevention trial. *Cancer Res.*, 50: 1216-1219, 1990.
  53. Cantaboni, A., Pezzotta, M. G., Sironi, M., and Porcellati, M. Quality assurance in pathology: cytologic and histologic correlation. *Acta Cytol.*, 36: 717-721, 1992.
  54. Holiday, D. B., McLary, J. W., Farley, M. L., Mabry, L. C., Cozens, D., Roby, T., Waldron, E., Underwood, R. D., Anderson, E., and Culbreth, W. Sputum cytology within and across laboratories: A reliability study. *Acta Cytol.*, 39: 195-206, 1995.
  55. Risse, E. K., Vooijs, G. P., and van't Hoff, M. A. Relationship between the cellular composition of sputum and the cytologic diagnosis of lung cancer. *Acta Cytol.*, 31: 170-176, 1987.
  56. MacDougall, B., and Weinerman, B. The value of sputum cytology. *J. Gen. Intern. Med.*, 7: 11-12, 1992.
  57. Tockman, M. S., Erozan, Y. S., Gupta, P., Piantadosi, S., Mulshine, J. L., and Ruckdeschel, J. C. The early detection of second primary lung cancers by sputum immunostaining. *Chest*, 106 (Suppl.): 385S-390S, 1994.
  58. Tockman, M. S., Gupta, P. K., Myers, J. D., Frost, J. K., Baylin, S. B., Gold, E. B., Chase, A. M., Wilkinson, P. H., and Mulshine, J. L. Sensitive and specific monoclonal antibody recognition of human lung cancer antigen on preserved sputum cells: a new approach to early lung cancer detection. *J. Clin. Oncol.*, 6: 1685-1693, 1988.
  59. Zhou, J., Mulshine, J. L., Unsworth, E. J., Scott, F. M., Avis, I. M., Vos, M. D., and Treason, A. M. Purification and characterization of a protein that permits early detection of lung cancer. *J. Biol. Chem.*, 271: 10760-10766, 1996.
  60. Qiao, Y. L., Tockman, M. S., Li, L., Erozan, Y. S., Yao, S. X., Barrett, M. J., Zhou, W. H., Giffen, C. A., Luo, X. C., and Taylor, P. R. A case-cohort study of an early biomarker of lung cancer in a screening cohort of Yunnan tin miners in China. *Cancer Epidemiol. Biomark. Prev.*, 6: 893-900, 1997.
  61. Grover, F. L., and Piantadosi, S. Recurrence and survival following resection of bronchioloalveolar carcinoma of the lung: the Lung Cancer Study Group experience. *Ann. Surg.*, 209: 779-790, 1989.
  62. Tockman, M. S. Advances in sputum analysis for screening and early detection of lung cancer. *Cancer Control*, 7: 19-24, 2000.
  63. Sueoka, E., Goto, Y., Sueoka, N., Kai, Y., Koza, T., and Fujiki, H. Heterogeneous nuclear ribonucleoprotein B1 as a new marker of early detection for human lung cancers. *Cancer Res.*, 59: 1404-1407, 1999.
  64. Mulshine, J. L. Reducing lung cancer risk. Early detection. *Chest*, 116: 493S-496S, 1999.
  65. Mao, L., Hruban, R. H., Boyle, J. O., Tockman, M., and Sidransky, D. Detection of oncogene mutations in sputum precedes diagnosis of lung cancer. *Cancer Res.*, 54: 1634-1637, 1994.
  66. Ruppert, J. M., Tokino, K., and Sidransky, D. Evidence for two bladder cancer suppressor loci on chromosome 9. *Cancer Res.*, 53: 5093-5095, 1993.
  67. Nawroz, H., van der Riet, P., Hruban, R. H., Koch, W., Ruppert, J. M., and Sidransky, D. Allelotype of head and neck squamous cell carcinoma. *Cancer Res.*, 54: 1152-1155, 1994.
  68. Miozzo, M., Sozzi, G., Musso, K., Filotti, S., Incabone, M., Pastorino, U., and Pierotti, M. A. Microsatellite alterations in bronchial and sputum specimens of lung cancer patients. *Cancer Res.*, 56: 2285-2288, 1996.
  69. Shapiro, G. I., Park, J. E., Edwards, C. D., Mao, L., Merlo, A., Sidransky, D., Ewen, M. E., and Rollins, B. J. Multiple mechanisms of p16<sup>INK4a</sup> inactivation in non-small cell lung cancer cell lines. *Cancer Res.*, 55: 6200-6209, 1995.
  70. Hamada, K., Kohno, T., Kawanishi, M., Ohwada, S., and Yokota, J. Association of CDKN2A(p16)/CDKN2B(p15) alterations and homozygous chromosome arm 9p deletions in human lung carcinoma. *Genes Chromosomes Cancer*, 22: 232-240, 1998.
  71. Belinsky, S. A., Nikula, K. J., Palmisano, W. A., Michels, R., Saccomanno, G., Gabrielson, E., Baylin, S. B., and Herman, J. G. Aberrant methylation of p16<sup>INK4a</sup> is an early event in lung cancer and a potential biomarker for early diagnosis. *Proc. Natl. Acad. Sci. USA*, 95: 11891-11896, 1998.
  72. Esteller, M., Sanchez-Cespedes, M., Rosell, R., Sidransky, D., Baylin, S. B., and Herman, J. G. Detection of aberrant promoter hypermethylation of tumor suppressor genes in serum DNA from non-small cell lung cancer patients. *Cancer Res.*, 59: 67-70, 1999.
  73. Nieburgs, H. E. Recent progress in the interpretation of malignancy associated changes (MAC). *Acta Cytol.*, 12: 445-453, 1968.
  74. Payne, P. W., Sebo, T. J., Doudkine, A., Garner, D., MacAulay, C., Lam, S., LeRiche, J. C., and Palcic, B. Sputum screening by quantitative microscopy: a reexamination of a portion of the National Cancer Institute Cooperative Early Lung Cancer Study. *Mayo Clin. Proc.*, 72: 697-704, 1997.
  75. Lam, S., Palcic, B., Garner, D., Beveridge, J., MacAulay, C., LeRiche, J., and Goldman, A. Lung Cancer Control Strategy in the New Millennium. *Lung Cancer*, 29 (Suppl. 2): 145, 2000.
  76. Ahrendt, S. A., Chow, J. T., Xu, L. H., Yang, S. C., Eisenberger, C. F., Esteller, M., Herman, J. G., Wu, L., Decker, P. A., Jen, J., and Sidransky, D. Molecular detection of tumor cells in bronchoalveolar lavage fluid from patients with early stage lung cancer. *J. Natl. Cancer Inst.*, 91: 332-339, 1999.
  77. Gazdar, A. F., and Minna, J. D. Molecular detection of early lung cancer. *J. Natl. Cancer Inst.*, 91: 299-301, 1999.
  78. Field, J. K., Liloglou, T., Xinarianos, G., Prime, W., Fielding, P., Walkshaw, M. J., and Turnbull, L. Genetic alterations in bronchial lavage as a potential marker for individuals with a high risk of developing lung cancer. *Cancer Res.*, 59: 2690-2695, 1999.
  79. Liloglou, T., Maloney, P., Xinarianos, G., Fear, S., and Field, J. K. Sensitivity and limitations of high throughput fluorescent microsatellite analysis for the detection of allelic imbalance. Application in lung tumors. *Int. J. Oncol.*, 16: 5-14, 2000.
  80. Fielding, P., Turnbull, L., Prime, W., Walkshaw, M., and Field, J. K. Heterogeneous nuclear ribonucleoprotein A2/B1 up-regulation in bronchial lavage specimens: a clinical marker of early lung cancer detection. *Clin. Cancer Res.*, 5: 404S-405S, 1999.
  81. Pantel, K., Schlimok, G., Angstwurm, M., Weckermann, D., Schmaus, W., Gath, H., Passlick, B., Izbicke, J. R., and Riethmüller, G. Methodological analysis of immunocytochemical screening for disseminated tumor cells in sputum. *Cancer*, 81: 100-108, 1993.

- inated epithelial tumor cells in bone marrow. *J. Hematother.*, 3: 165-173, 1994.
82. Lambrechts, A. C., van't Veer, L. J., and Rodenhuis, S. The detection of minimal numbers of contaminating epithelial tumor cells in blood or bone marrow: use, limitations and future of RNA-based methods. *Ann. Oncol.*, 9: 1269-1276, 1998.
83. Steinman, C. R. Free DNA in serum and plasma from normal adults. *J. Clin. Investig.*, 66: 1391-1399, 1980.
84. Raptis, L., and Menard, H. A. Quantitation and characterization of plasma DNA in normals and patients with lupus erythematosus. *J. Clin. Investig.*, 66: 1391-1399, 1980.
85. Leon, S. A., Shapiro, B., Sklaroff, D. M., and Yaros, M. J. Free DNA in the serum of cancer patients and the effect of therapy. *Cancer Res.*, 37: 646-650, 1977.
86. Stroun, M., Anker, P., Maurice, P., Lyautey, J., Lederrey, C., and Beljanski, M. Neoplastic characteristics of the DNA found in the plasma of cancer patients. *Oncology*, 46: 318-322, 1989.
87. Shapiro, B., Chakrabaty, M., Cohn, E., and Leon, S. A. Determination of circulating DNA levels in patients with benign or malignant gastrointestinal disease. *Cancer (Phila.)*, 51: 2116-2120, 1983.
88. Chen, X. Q., Stroun, M., Magnenat, J.-L., Nicod, L. P., Kurt, A. M., Lyautey, J., Lederrey, C., and Anker, P. Microsatellite alterations in plasma DNA of small cell lung cancer patients. *Nat. Med.*, 2: 1033-1037, 1996.
89. Anker, P., Lefort, F., Vasioukhin, V., Lyautey, J., Lederrey, C., Chen, X. Q., Stroun, M., Mulcahy, H. E., and Farthing, M. J. K-ras mutations are found in DNA extracted from the plasma of colorectal cancer patients. *Gastroenterology*, 112: 1114-1129, 1997.
90. Kopreski, M. S., Benko, F. A., Kwee, C., Leitzel, K. E., Eskander, E., Lipton, A., and Gocke, C. D. Detection of mutant K-ras DNA in plasma or serum of patients with colorectal cancer. *Br. J. Cancer*, 76: 1293-1299, 1997.
91. Hibi, K., Robinson, R., Wu, L., Hamilton, S. R., Sidransky, D., and Jen, J. Molecular detection of genetic alterations in the serum of colorectal cancer patients. *Cancer Res.*, 58: 1405-1407, 1998.
92. Sorenson, G. D., Pribish, D. M., Valone, F. H., Memoli, V. A., Bizik, D. J., and Yao, S. L. Soluble normal and mutated DNA sequences from single copy genes in human blood. *Cancer Epidemiol. Biomark. Prev.*, 3: 67-71, 1994.
93. Mulcahy, H. E., Lyautey, J., Lederrey, C., Chen, X., Anker, P., Alstead, E. N., Ballinger, A., Farthing, M. J., and Stroun, M. A prospective study of K-ras mutations in the plasma of pancreatic cancer patients. *Clin. Cancer Res.*, 4: 271-275, 1998.
94. Vasioukhin, V., Anker, P., Maurice, P., Lyautey, J., Lederrey, C., and Stroun, M. Point mutations of the N-ras in the blood plasma DNA of patients with myelodysplastic syndrome or acute myelogenous leukemia. *Br. J. Haematol.*, 86: 774-779, 1994.
95. Woolner, L. B. Pathology of cancer detected cytologically. In: *Atlas of Early Lung Cancer*. National Cancer Institute, NIH, United States Department of Health and Human Services, pp. 107-203. Tokyo: Igaku-Shoin, 1983.
96. Kato, H., and Cortese, D. A. Early detection of lung cancer by means of hematoporphyrin derivative fluorescence and laser photoradiation. *Clin. Chest Med.*, 6: 237-253, 1985.
97. Profio, A. E., Dörion, D. R., and King, E. G. L. Laser fluorescence bronchoscopy for localization of occult lung tumors. *Med. Phys. (NY)*, 6: 532-535, 1979.
98. Kinsey, J. H., and Cortese, D. A. Endoscopic system for simultaneous visual examination and electronic detection of fluorescence. *Rev. Sci. Instrum.*, 51: 1402-1406, 1980.
99. Profio, A. E., Dörion, D. R., and Sarnak, J. Fluorometer for endoscopic diagnosis of tumors. *Med. Phys. (NY)*, 11: 516-520, 1984.
100. Montan, S., Svanberg, K., and Svanberg, S. Multicolor imaging and contrast enhancement in cancer-tumor localization using laser-induced fluorescence in hematoporphyrin-derivative-bearing tissue. *Optics Lett.*, 10: 56-58, 1985.
101. Kato, H., Imaizumi, T., Aizawa, K., Iwabuchi, H., Yamamoto, H., Ikeda, N., Tsuchida, T., Tamachi, Y., Ito, T., and Hayata, Y. Photodynamic diagnosis in respiratory tract malignancy, using an excimer dye laser system. *J. Photochem. Photobiol. Biol.*, 6: 189-196, 1990.
102. Kato, H., and Ikeda, N. The role of fluorescence diagnosis in the early detection of high-risk bronchial lesions. *J. Bronchol.*, 5: 273-274, 1998.
103. Falcic, B., Lam, S., Hung, J., and MacAulay, C. Detection and localization of early lung cancer by imaging techniques. *Chest*, 99: 742-743, 1991.
104. Hung, J., Lam, S., LeRiche, J. C., and Falcic, B. Autofluorescence of normal and malignant bronchial tissue. *Lasers Surg. Med.*, 11: 99-105, 1991.
105. Lam, S., Kennedy, T., Unger, M., Miller, Y. E., Gelmon, D., Rusch, V., Gipe, B., Howard, D., LeRiche, J. C., Goldman, A., and Gazdar, A. F. Localization of bronchial intraepithelial neoplastic lesions by fluorescence bronchoscopy. *Chest*, 113: 696-702, 1998.
106. Kuric, J. M., Lee, J. S., Morice, R. C., Walsh, G. L., Khuri, F. R., Broxson, A., Ro JY, Franklin, W. A., Yu, R., and Hong, W. K. Autofluorescence bronchoscopy in the detection of squamous metaplasia and dysplasia in current and former smokers. *J. Natl. Cancer Inst.*, 90: 991-995, 1998.
107. Venmans, B. J., van der Linden, J. C., van Boxem, A. J., Postmus, P. E., Smit, E. F., and Sutedja, G. Early detection of pre-invasive lesions in high risk patients. A comparison of conventional fiberoptic and fluorescence bronchoscopy. *J. Bronchol.*, 5: 280-283, 1998.
108. Vermeylen, P., Pierard, P., Roufosse, C., Bosschaerts, T., Verhest, A., Sculier, J.-P., and Ninane, V. Detection of bronchial preneoplastic lesions and early lung cancer with fluorescence bronchoscopy: a study about its ambulatory feasibility under local anaesthesia. *Lung Cancer (Limerick)*, 25: 161-168, 1999.
109. Kennedy, T., Hirsch, F. R., Miller, Y., Prindiville, S., Bunn, P. A., Jr., and Franklin, W. A randomized study of fluorescence bronchoscopy versus white-light bronchoscopy for early detection of lung cancer in high risk patients. *Lung Cancer (Limerick)*, 29 (Suppl. 2): 244, 2000.
110. Gazdar, A. F., and Minna, J. D. Angiogenesis and the multistage development of lung cancers. *Clin. Cancer Res.*, 6: 1611-1612, 2000.
111. Henschke, C. J., Miettinen, O. S., Yankellevitz, D. F., Libby, D. M., and Smith, J. P. Radiographic screening for cancer: proposed paradigm for requisite research. *Clin. Imaging*, 18: 6-20, 1994.
112. Miettinen, O. S. Screening for lung cancer. *Radiol. Clin. N. Am.*, 38: 479-486, 2000.
113. Marcus, P. M., Bergstralh, E. J., Fagerstrom, M., Williams, D. E., Fontana, R., Taylor, W. F., and Prorok, P. C. Lung cancer mortality in the Mayo Lung Project: impact of extended follow-up. *J. Natl. Cancer Inst.*, 92: 1308-1316, 2000.
114. Black, W. C. Overdiagnosis. An underrecognized cause of confusion and harm in cancer screening. *J. Natl. Cancer Inst.*, 92: 1280-1282, 2000.
115. The Health, and Welfare Statistics Foundation. The current issue on the screening project for elderly people by the Ministry of Health and Welfare Japan. *J. Health Welfare*, 42: 37, 1995.
116. Sobue, T., Suzuki, T., and Naruke, T. The Japanese Lung-Cancer-Screening Research Group: a case-control study for evaluating lung cancer screening in Japan. *Int. J. Cancer*, 50: 230-237, 1992.
117. Okamoto, N., Suzuki, T., Hasegawa, H., Gotoh, T., Hagiwara, S., Sekimoto, M., and Kaneko, M. Evaluation of a clinic-based screening program for lung cancer with a case-control design in Kanagawa, Japan. *Lung Cancer (Limerick)*, 25: 77-85, 1999.
118. Muhm, J. R., Miller, W. F., Fontana, R. S., Sanderson, D. R., and Uhlenhuth, M. A. Lung cancer detected during a screening program using four-month chest radiographs. *Radiology*, 148: 609-615, 1983.
119. Hayabuchi, N., Russell, W. J., and Murakami, J. Problems in radiographic detection and diagnosis of lung cancer. *Acta Radiol. (CPI)*, 30: 163-167, 1989.
120. Woodring, J. H. Pitfalls in the radiologic diagnosis of lung cancer. *Am. J. Roentgenol.*, 154: 1165-1175, 1990.



121. Greene, R. E. Missed lung nodules: lost opportunities for cancer care. *Radiology*, 182: 8-9, 1992.
122. Austin, J. H. M., Romney, B. M., Goldsmith, L. S. Missed bronchogenic carcinoma: radiographic findings in 27 patients with a potentially resectable lesions evident in retrospect. *Radiology*, 182: 115-122, 1992.
123. Schaner, E. G., Chang, A. E., Doppan, J. L., Conkle, D. M., Flye, M. W., and Rosenberg, S. A. Comparison of computed and conventional whole lung tomography in detecting pulmonary nodules. *Am. J. Roentgenol.*, 131: 51-54, 1978.
124. Webb, W. R. Radiologic evaluation of solitary pulmonary nodule. *Am. J. Roentgenol.*, 154: 701-708, 1990.
125. Kalender, W. A., Seissler, W., Klotz, E., and Vock, P. Spiral volumetric CT with single-breath-hold technique, continuous transport, and continuous scanner rotation. *Radiology*, 176: 181-183, 1990.
126. Costello, P., Anderson, W. A., and Blume, D. Pulmonary nodule: evaluation with spiral volumetric CT. *Radiology*, 179: 875-876, 1991.
127. Remy-Jardin, M., Giraud, F., and Marquette, C. H. Pulmonary nodules: detection with thick section spiral CT versus conventional CT. *Radiology*, 187: 513-520, 1993.
128. Davis, S. CT evaluation for pulmonary metastases in patients with extrathoracic malignancy. *Radiology*, 180: 1-12, 1991.
129. Kaneko, M., Eguchi, K., Ohmatsu, H., Kakinuma, R., Naruke, T., Suemasu, K., and Moriyama, N. Peripheral lung cancer: screening and detection with low-dose spiral CT versus radiography. *Radiology*, 201: 798-802, 1996.
130. Sone, S., Takashima, S., Li, F., Yang, Z., Honda, T., Maruyama, Y., Hasegawa, M., Yamada, T., Kubo, K., Hanamura, K., and Asakura, K. Mass screening for lung cancer with mobile spiral computed tomography scanner. *Lancet*, 351: 1242-1245, 1998.
131. Henschke, C. J., McCauley, D. J., Yankelevitz, D. F., Naidich, D. P., McGuinness, G., Miettinen, O. S., Libby, D. M., Pasmantier, M. W., Koizumi, J., Altorki, N. K., and Smith, J. P. Early Lung Cancer Action Project: overall design and findings from baseline screening. *Lancet*, 354: 99-105, 1999.
132. Kakinuma, R., Ohmatsu, H., Kaneko, M., Eguchi, K., Naruke, T., Nagai, K., Nishiwaki, Y., Suzuki, A., and Moriyama, N. Detection failures in spiral CT screening for lung cancer: analysis of CT findings. *Radiology*, 212: 61-66, 1999.
133. Paranjpe, D. V., and Bergin, C. J. Spiral CT of the lungs: optimal technique and resolution compared with conventional CT. *Am. J. Roentgenol.*, 162: 561-567, 1994.
134. Seltzer, S. E., Judy, P. F., Adams, D. F., Jacobson, F. L., Stark, P., Kikinis, R., Swenson, R. G., Hooton, S., Head, B., and Feldman, U. Spiral CT of the chest: comparison of cine and film-based viewing. *Radiology*, 197: 73-78, 1995.
135. Croisille, P., Souto, M., Cova, M., Wood, S., Afework, Y., Kuhlman, J. E., and Zerhouni, E. A. Pulmonary nodules: improved detection with vascular segmentation and extraction with spiral CT. *Radiology*, 197: 397-401, 1995.
136. Kanazawa, K., Kawato, Y., Niki, N., Satoh, H., Ohmatsu, H., Kakinuma, R., and Kaneko, M. Computer-aided diagnosis for pulmonary nodules based on helical CT images. *Comput. Med. Imaging Graph.*, 22: 157-167, 1998.
137. Buckner, N., Agnello, K., Yankelevitz, D. F., Mancuso, C., and Henschke, C. J. Smoking habits and overall satisfaction after early lung cancer screening using low-dose CT. *ALA/ATS International Conference*, 1999.
138. Coleman, R. E. PET in lung cancer. *J. Nucl. Med.*, 40: 814-820, 1999.
139. Maina, A., Wistuba, I. I., Virmani, A. K., Sakaguchi, M., Park, J., Sucky, A., Milchgrub, S., Gibbons, D., Minna, J. D., and Gazdar, A. F. Enrichment of epithelial cells for molecular studies. *Nat. Med.*, 5: 459-463, 1999.
140. Kopelovich, L., Henson, D. E., Gazdar, A. F., Dubb, B., Srivastava, S., Kelloff, G. J., and Greenwald, P. Surrogate anatomic/functional sites for evaluating cancer risk: an extension of the field effect. *Clin. Cancer Res.*, 5: 3899-3905, 1999.
141. Crowell, R. E., Gilliland, F. D., Temes, R. T., Harms, H. J., Neft, R. E., Heaphy, E., Auckley, D. H., Crooks, L. A., Jordan, S. W., Samet, J. M., Lechner, J. F., and Belinsky, S. A. Detection of trisomy 7 in nonmalignant bronchial epithelium from lung cancer patients and individuals at risk for lung cancer. *Cancer Epidemiol. Biomark. Prev.*, 5: 631-637, 1996.
142. Somers, V. A., van Henten, A. M., ten Velde, G. P., Arends, J. W., and Thunnissen, F. B. Additional value of K-ras point mutations in bronchial wash fluids for diagnosis of peripheral lung tumours. *Eur. Respir. J.*, 13: 1120-1124, 1999.
143. Yabata, N., Ohyashiki, K., Ohyashiki, J. H., Iwama, H., Hayashi, S., Ando, K., Hirano, T., Tsuchida, T., Kato, H., Shay, J. W., and Toyama, K. Telomerase activity in lung cancer patients and individuals at risk for lung cancer. *Cancer Epidemiol. Biomark. Prev.*, 5: 631-637, 1996.
144. Wistuba, I. I., and Gazdar, A. F. Molecular abnormalities in the sequential development of lung carcinoma. In: S. Srivastava, D. E. Henson, and A. F. Gazdar (eds.), *Molecular Pathology of Early Cancer*, pp. 265-276. Amsterdam: IOS Press, 1999.
145. Yashima, K., Litzky, L. A., Kaiser, L., Rogers, T., Lam, S., Wistuba, I. I., Milchgrub, S., Srivastava, S., Piatyszek, M. A., Shay, J. W., and Gazdar, A. F. Telomerase expression in respiratory epithelium during the multistage pathogenesis of lung carcinomas. *Cancer Res.*, 57: 2373-2377, 1997.
146. Scott, F. M., Treston, A. M., Shaw, G. L., Avis, I., Sorenson, J., Kelly, K., Dempsey, E. C., Cantor, A. B., Tockman, M., and Mulshine, J. L. Peptide amidating activity in human bronchoalveolar lavage fluid. *Lung Cancer (Limerick)*, 14: 239-251, 1996.
147. Scott, F. M., Modali, R., Lehman, T. A., Seddon, M., Kelly, K., Dempsey, E. C., Wilson, V., Tockman, M. S., and Mulshine, J. L. High frequency of K-ras codon 12 mutations in bronchoalveolar lavage fluid of patients at high risk for second primary lung cancer. *Clin. Cancer Res.*, 3: 479-482, 1997.
148. Mills, N. E., Fishman, C. L., Rom, W. N., and Jacobson, D. R. *Ras* oncogene detection in bronchoalveolar lavage fluid from patients with lung cancer. *Lung Cancer (Limerick)*, 1 (Suppl.): 11, 1994.
149. Nawroz, H., Koch, W., Anker, P., Stroun, M., and Sidransky, D. Microsatellite alterations in serum DNA of head and neck cancer patients. *Nat. Med.*, 2: 1035-1037, 1996.
150. Brennan, J. A., Mao, L., Hruban, R. H., Boyle, J. O., Eby, Y. J., Koch, W. M., Goodman, S. N., and Sidransky, D. Molecular assessment of histopathological staging in squamous-cell carcinoma of the head and neck. *N. Engl. J. Med.*, 332: 429-435, 1995.
151. Ahrendt, S. A., Yang, S. C., Wu, L., Westra, W. H., Jen, J., Califano, J. A., and Sidransky, D. Comparison of oncogene mutation detection and telomerase activity for the molecular staging of non-small cell lung cancer. *Clin. Cancer Res.*, 3: 1207-1214, 1997.
152. Gazdar, A. F., and Minna, J. D. NCI series of cell lines: an historical perspective. *J. Cell. Biochem.*, 24 (Suppl.): 1-11, 1996.
153. Wistuba, I. I., Bryant, D., Behrens, C., Milchgrub, S., Virmani, A. K., Ashfaq, R., Minna, J. D., and Gazdar, A. F. Comparison of features of human lung cancer cell lines and their corresponding tumors. *Clin. Cancer Res.*, 5: 991-1000, 1999.
154. Amstad, P., Reddel, R. R., Pfeifer, A., Malan, S. L., Mark, G. D., and Harris, C. C. Neoplastic transformation of a human bronchial epithelial cell line by a recombinant retrovirus encoding viral Harvey *ras*. *Mol. Carcinog.*, 1: 151-160, 1988.
155. Franklin, W. A., Folkvord, J. M., Varella-Garcia, M., Kennedy, T., Proudfoot, S., Cook, R., Dempsey, E. C., Helm, K., Burn, P. A., and Miller, Y. E. Expansion of bronchial epithelial cell populations by *in vitro* culture of explants from dysplastic and histologically normal sites. *Am. J. Respir. Cell Mol. Biol.*, 15: 297-304, 1996.
156. Kelsey, K., Spitz, M., Zuo, A., and Wiencke, J. Deletion of glutathione S-transferase class  $\mu$  and class  $\theta$  genes interacts to enhance susceptibility to lung cancer in minority populations. *Cancer Causes Control*, 1997.

157. Wu, X., Zhao, Y., Honn, S. E., Tomlinson, G. E., Minna, J. D., Hong, W. K., and Spitz, M. R. Benzo[a]pyrene diol epoxide-induced 3p21.3 aberrations and genetic predisposition to lung cancer. *Cancer Res.*, 58: 1605-1608, 1998.
158. Virmani, A. K., Fong, K. M., Kodagoda, D., McIntire, D., Hung, J., Tonk, V., Minna, J. D., and Gazdar, A. F. Allelotyping demonstrates common and distinct patterns of chromosomal loss in human lung cancer types. *Genes Chromosomes Cancer*, 21: 308-319, 1998.
159. Takahashi, T., Nau, M. M., Chiba, I., Birrer, M. J., Rosenberg, R. K., Vinocour, M., Levitt, M., Pass, H., Gazdar, A. F., and Minna, J. D. p53: a frequent target for genetic abnormalities in lung cancer. *Science (Washington DC)*, 246: 491-494, 1989.
160. Forgacs, E., Biesterveld, E. J., Sekido, Y., Fong, K. M., Muneer, S., Wistuba, I., Milchgrub, S., Breznitschek, R., Virmani, A., Gazdar, A. F., and Minna, J. D. Mutation analysis of the *PTEN/MMAC1* gene in lung cancer. *Oncogene*, 17: 1557-1565, 1998.
161. Mitsudomi, T., Steinberg, S., Oie, H. K., Mulshine, J. L., Phelps, R., Viallet, J., Pass, H., Minna, J. D., and Gazdar, A. F. *ras* gene mutations in non-small cell lung cancers are associated with shortened survival irrespective of treatment intent. *Cancer Res.*, 51: 4999-5002, 1991.
162. Sozzi, G., Veronese, M. L., Negrini, M., Baffa, R., Corticelli, M. G., Inoue, H., Tonielli, S., Pilotti, S., Ohta, M., Huebner, K., and Croce, C. M. The *FHIT* gene at 3p14.2 is abnormal in lung cancer. *Cell*, 85: 17-26, 1996.
163. Ambazhagan, R., Tihan, T., Bornman, D. M., Johnston, J. C., Saltz, J. H., Weigering, A., Piantadosi, S., and Gabrielson, E. Classification of small cell lung cancer and pulmonary carcinoid by gene expression profiles. *Cancer Res.*, 59: 5119-5122, 1999.
164. Hibi, K., Liu, Q., Beaudry, G. A., Madden, S. L., Westra, W. H., Wehage, S. L., Yang, S. C., Heimiller, R. F., Bertelsen, A. H., Sidransky, D., and Jen, J. Serial analysis of gene expression in non-small cell lung cancer. *Cancer Res.*, 58: 5690-5694, 1998.
165. Sozzi, G., Pastorino, U., Moiraghi, L., Tagliabue, E., Pezzella, F., Ghirelli, C., Tonielli, S., Sard, L., Huebner, K., Pierotti, M. A., Croce, C. M., and Pilotti, S. Loss of *FHIT* function in lung cancer and preinvasive bronchial lesions. *Cancer Res.*, 58: 5032-5037, 1998.
166. Tockman, M. S., and Mulshine, J. L. Sputum screening by quantitative microscopy: a new dawn for detection of lung cancer? *Mayo Clin. Proc.*, 72: 788-790, 1997.
167. Jzso, J., and Hittelman, W. N. Characterization of multistep tumorigenesis by *in situ* hybridization. In: M. Andreeff and D. Pinkel (eds.), *Introduction to FISH*. In press, 2001.
168. Neft, R. E., Crowell, R. E., Gilliland, F. D., Murphy, M. M., Lane, J. L., Harms, H., Coons, T., Heaphy, E., Belinsky, S. A., and Lechner, J. F. Frequency of trisomy 20 in nonmalignant bronchial epithelium from lung cancer patients and cancer-free former uranium miners and smokers. *Cancer Epidemiol. Biomark. Prev.*, 7: 1051-1054, 1998.
169. Heppell-Parton, A. C., Nacheva, E., Carter, N. P., and Rabbitts, P. H. A combined approach of conventional and molecular cytogenetics for detailed karyotypic analysis of the small cell lung carcinoma cell line U2020. *Cancer Genet. Cytogenet.*, 108: 110-119, 1999.
170. Lechner, J. F., Neft, R., Gilliland, F. D., Crowell, R. E., Auckely, D. H., Ternes, R. T., and Belinsky, S. A. Individuals at high risk for lung cancer have airway epithelial cells with chromosome aberrations frequently found in lung tumor cells. *In Vivo* 12: 23-26, 1998.
171. Walch, A. K., Zitzelsberger, H. F., Aubele, M. M., Maitis, A. E., Bauchinger, M., Candidus, S., Prauer, H. W., Werner, M., and Hofer, H. Typical and atypical carcinoid tumors of the lung are characterized by 11q deletions as detected by comparative genomic hybridization. *Am. J. Pathol.*, 153: 1089-1098, 1998.
172. Petersen, I., Bujard, M., Petersen, S., Wolf, G., Goeze, A., Schwendel, A., Langreck, H., Gellert, K., Reichel, M., Just, K., du Manoir, S., Cremer, T., Dietel, M., and Ried, T. Patterns of chromosomal imbalances in adenocarcinoma and squamous cell carcinoma of the lung. *Cancer Res.*, 57: 2331-2335, 1997.
173. Smith, A. L., Hung, J., Walker, L., Rogers, T. E., Vuitch, F., Lee, E., and Gazdar, A. F. Extensive areas of aneuploidy are present in the respiratory epithelium of lung cancer patients. *Br. J. Cancer*, 73: 203-209, 1996.
174. MacAulay, C. E., Lam, S., Klein-Parker, H., Gazdar, A., Guillaud, M., Payne, P., LeRiche, J., Dawe, C., Band, P., and Polcic, B. Intermediate endpoint biomarkers for lung cancer chemoprevention. *SPICE*, 3260: 207-211, 1998.



**Localization of tissue inhibitor of metalloproteinases 1 (TIMP-1) in human colorectal adenoma and adenocarcinoma.**

Holten-Andersen MN, Hansen U, Brunner N, Nielsen HJ, Illemann M, Nielsen BS.

The Finsen Laboratory, Rigshospitalet, Denmark.

Tissue inhibitor of matrix metalloproteinases 1 (TIMP-1) inhibits the proteolytic activity of matrix metalloproteinases and hereby prevents cancer invasion. However, TIMP-1 also possesses other functions such as inhibition of apoptosis, induction of malignant transformation and stimulation of cell-growth. We have previously demonstrated that TIMP-1 is elevated in blood from colorectal cancer patients and that high TIMP-1 levels predict poor prognosis. To clarify the role of TIMP-1 in colorectal tumorigenesis, the expression pattern of TIMP-1 in benign and malignant colorectal tumors was studied. In all of 24 cases of colorectal adenocarcinoma TIMP-1 mRNA was detected by in situ hybridization. In all cases TIMP-1 expression was found in fibroblast-like cells located at the invasive front but was seen only sporadically in normal mucosa. No TIMP-1 mRNA was seen in any of the cases in benign or malignant epithelial cells, in vascular cells or smooth muscle cells. Comparison of sections processed for TIMP-1 in situ hybridization with sections immunohistochemically stained with antibodies against TIMP-1 showed good correlation between TIMP-1 mRNA and immunoreactivity. Combining TIMP-1 in situ hybridization with immunohistochemical staining for alpha-smooth muscle actin or CD68 showed TIMP-1 mRNA in myofibroblasts but not in macrophages. TIMP-1 mRNA was detected in 2 of 7 adenomatous polyps in the adenoma area: in both cases associated with focal stromal inflammation at the epithelial-stromal interface. In conclusion, TIMP-1 expression is a rare event in benign human colon tissue but is highly expressed by myofibroblasts in association with invading colon cancer cells.

PMID: 15386409 [PubMed - indexed for MEDLINE]

## Localization of Tissue Inhibitor of Metalloproteinases 1 (TIMP-1) in Human Colorectal Adenoma and Adenocarcinoma

Mads N. Holten-Andersen<sup>1,2\*</sup>, Ulla Hansen<sup>3</sup>, Nils Brünnér<sup>2</sup>, Hans Jørgen Nielsen<sup>4</sup>, Martin Blemann<sup>1</sup> and Boye Schnack Nielsen<sup>1</sup>

<sup>1</sup>The Finsen-Laboratory, Rigshospitalet, Denmark

<sup>2</sup>The Institute for Pharmacology and Pathobiology, The Royal Veterinary and Agricultural University, Denmark

<sup>3</sup>Department of Pathology, University Hospital Hvidovre, Denmark

<sup>4</sup>Department of Surgical Gastroenterology, University Hospital Hvidovre, Denmark

Tissue inhibitor of matrix metalloproteinases 1 (TIMP-1) inhibits the proteolytic activity of matrix metalloproteinases and hereby prevents cancer invasion. However, TIMP-1 also possesses other functions such as inhibition of apoptosis, induction of malignant transformation and stimulation of cell-growth. We have previously demonstrated that TIMP-1 is elevated in blood from colorectal cancer patients and that high TIMP-1 levels predict poor prognosis. To clarify the role of TIMP-1 in colorectal tumorigenesis, the expression pattern of TIMP-1 in benign and malignant colorectal tumors was studied. In all of 24 cases of colorectal adenocarcinoma TIMP-1 mRNA was detected by *in situ* hybridization. In all cases TIMP-1 expression was found in fibroblast-like cells located at the invasive front but was seen only sporadically in normal mucosa. No TIMP-1 mRNA was seen in any of the cases in benign or malignant epithelial cells, in vascular cells or smooth muscle cells. Comparison of sections processed for TIMP-1 *in situ* hybridization with sections immunohistochemically stained with antibodies against TIMP-1 showed good correlation between TIMP-1 mRNA and immunoreactivity. Combining TIMP-1 *in situ* hybridization with immunohistochemical staining for  $\alpha$ -smooth muscle actin or CD68 showed TIMP-1 mRNA in myofibroblasts but not in macrophages. TIMP-1 mRNA was detected in 2 of 7 adenomatous polyps in the adenoma area: in both cases associated with focal stromal inflammation at the epithelial-stromal interface. In conclusion, TIMP-1 expression is a rare event in benign human colon tissue but is highly expressed by myofibroblasts in association with invading colon cancer cells.

**Key words:** TIMP-1; MMP; *in situ* hybridization; immunohistochemistry; myofibroblast

A prerequisite for cancer cell invasion and metastasis is the breakdown of tissue barriers mediated by proteolytic enzymes such as the matrix metalloproteinases (MMP).<sup>1,2</sup> Under normal physiologic conditions, the tissue degrading activities of the MMPs are kept at bay by the presence of the naturally occurring inhibitors: tissue inhibitors of metalloproteinases (TIMP). TIMP-1, a 28 kDa glycoprotein demonstrated to be present in most bodily tissues and fluids, binds and inhibits MMPs in a 1:1 stoichiometric manner.<sup>3,4</sup> Overexpression of TIMP-1 in various cancer models has shown a suppressive role in the malignant progression.<sup>5</sup> However, as opposed to this anti-invasive role of TIMP-1, several recent studies have demonstrated quite different functions of this MMP-inhibitor including stimulation of cell growth, malignant transformation and inhibition of apoptosis, suggesting a possible tumor-promoting role of TIMP-1 in very early stages of tumorigenesis.<sup>6–10</sup> Thus, it has been speculated that TIMP-1 may actually play a dual role in cancer progression and metastasis.<sup>11</sup>

Several studies have demonstrated that tumor tissue levels of MMP mRNA and protein are significantly increased in various malignant diseases and that such MMP elevations are correlated with cancer cell invasion, metastasis and short patient survival.<sup>12,13</sup> In addition, many reports have described similar overexpression of TIMP-1 mRNA and protein in several cancer types.<sup>14–20</sup> Moreover, we and others have demonstrated that measurement of increased plasma levels of TIMP-1 by immunoassay serves as a strong marker for short survival and recurrence of disease in patients with colorectal cancer.<sup>21–23</sup> Similarly, a strong correlation

between high protein levels and poor prognosis is known for the type-1 plasminogen activator inhibitor (PAI-1).<sup>24,25</sup> Considering the protease inhibiting function of these inhibitors, these findings seemed controversial; however, alternative functions have been reported both for TIMP-1 as mentioned above as well as for PAI-1.<sup>26</sup>

In order to better understand the role of TIMP-1 in colorectal cancer, histochemical analyses may provide some indications. A number of studies of the localization of TIMP-1 in colorectal cancer have been published; however, the results of these reports are somewhat contradictory. Newell and colleagues<sup>27</sup> reported that TIMP-1 mRNA was expressed both in invasive adenocarcinoma, carcinoma *in situ* and adenoma and that the expression was observed in both the stromal as well as the epithelial compartment of the tissues studied. In contrast, Zeng and colleagues<sup>12,20</sup> reported that TIMP-1 mRNA was expressed only in the stromal compartment of colorectal adenocarcinomas in spindle-shaped cells surrounding the invasive cancer cells. The results of immunohistochemical studies of TIMP-1 in colon are also conflicting: Hewitt and colleagues<sup>19</sup> reported that TIMP-1 was expressed in the connective tissue and basement membrane in both normal mucosa, adenomas and adenocarcinomas with only little staining of the neoplastic epithelium. On the other hand, Tomita and colleagues<sup>28</sup> reported that TIMP-1 was expressed in both stromal and epithelial cells in colonic polyps and adenomas, as well as in adenocarcinomas, in which the neoplastic cells were strongly immunoreactive.

In order to resolve these inconsistencies, we undertook our study and by *in situ* hybridization and immunohistochemistry demonstrated that TIMP-1 is expressed in myofibroblasts in the stroma at the invasive front of colorectal adenocarcinomas. Because TIMP-1 was virtually absent from normal colorectal epithelium, we evaluated the possibility of using TIMP-1 as a diagnostic tool to differentiate colorectal adenomas from Dukes' stage A colorectal adenocarcinomas.

### Material and methods

#### Tissue samples

All tissue material included was obtained from University Hospital of Hvidovre (Copenhagen, Denmark) in accordance with a permission given by the local scientific ethical committee (KF 01-078/93). Fourteen archival samples (formalin fixed and paraffin embedded) collected from 1989 to 1993 included Dukes' stage A colorectal adenocarcinomas ( $n = 8$ ) and colorectal adenomatous polyps ( $n = 6$ , 3 were pedunculated (1 with mild and 2 with

Grant sponsor: The Danish Cancer Society; Grant sponsor: European commission; Grant number: QLGI-CT-2000-011131; Grant sponsor: Weimann Foundation.

\*Correspondence to: Mads N Holten-Andersen, The Institute for Pharmacology and Pathobiology, The Royal Veterinary and Agricultural University, Denmark.

Received 16 February 2004; Accepted after revision 7 June 2004  
DOI 10.1002/ijc.20566

Published online 13 September 2004 in Wiley InterScience (www.interscience.wiley.com).

moderate dysplasia) and 3 were sessile (1 with moderate and 2 with focally severe dysplasia). Samples from 16 colorectal adenocarcinomas (1 Duke's stage A, 6 Duke's stage B, 8 Duke's stage C and 1 Duke's stage D), 1 villous adenoma and 1 malignant colon lymphoma were prospectively collected during 1999–2000. These prospectively collected tissue specimens were dissected so that samples contained both normal mucosa and tumor tissue and were obtained within 30 min following surgical bowel resection. The specimens were immediately fixed in 4% neutral buffered formalin for 20–24 hr and then paraffin embedded. The 14 archival samples had also been formalin fixed and paraffin embedded.

#### Generation of nonoverlapping TIMP-1 cDNA fragments by PCR

The full length TIMP-1 cDNA (GenBank NM\_003254) cloned in pSP64 vector<sup>29</sup> was used as template to generate 2 nonoverlapping PCR fragments for *in vitro* transcription, and named f104 (bp 56–378) and f106 (bp 398–680). First, the whole insert (~780 bp) was cut out by digestion with *Hind*III and *Bam*HI and purified after agarose gel electrophoresis using the Qiaex II gel extraction kit (Qiagen, Crawley, United Kingdom). To generate nonoverlapping antisense probes and the corresponding sense probe, 2 PCR fragments were generated using upstream primers flanked by a linker sequence containing an *Eco*RI restriction enzyme site (underlined nucleotides) and a T3 polymerase binding sequence (boldface) 5'-(**gagatc**attcattaccctcactaaaggaga)-3', and downstream primers flanked by a linker sequence containing a *Bam*HI restriction enzyme site and a T7 polymerase binding sequence 5'-(**ggatc**ctataacgactactataggag)-3'. The TIMP-1 specific upstream primers were 5'-accaccatggccccccttg-3' for f104 and 5'-(linker)-gcaggatggactcttgcaca-3' for f106, and the downstream primers were 5'-(linker)-actctcgtcgggttg-3' for f104 and 5'-(linker)-tatctgggaccgaggag-3' for f106. PCR using the 2 f104 primers or the 2 f106 primers was done as previously described.<sup>30</sup>

The PCR products were purified by column chromatography using S-200HR microspin columns (Amersham Pharmacia Biotech, Inc., Piscataway, NJ), and their size tested by agarose gel electrophoresis. Both migrated as ~300 bp fragments in accordance with the predicted size (322 and 282 bp, respectively). An ABI PRISM 310 genetic analyzer was employed for DNA sequencing analysis and was performed according to the manufacturer's instructions (Perkin Elmer, Applied Biosystems, Foster City, CA) using the primers specified above. The DNA sequences obtained were confirmed by comparison with the specific TIMP-1 cDNA nucleic acid sequence (GenBank NM\_003254).

Plasmids containing human MMP-2 cDNA (pCol7201, bp 647–1284) and human MMP-9 cDNA (pCol9202, bp 1751–2326) have been described elsewhere.<sup>31</sup>

#### In vitro transcription

Antisense and sense riboprobes were labeled with <sup>35</sup>S UTP (NEN, Boston, MA) by *in vitro* transcription using T7 and T3 RNA polymerases (Roche, Basel, Switzerland). The DNA template was digested with DNase (Promega, Madison, WI). Nonincorporated <sup>35</sup>S UTP and DNA was removed by column chromatography using S-200HR microspin columns (Amersham Pharmacia Biotech, Inc., Piscataway, NJ). The <sup>35</sup>S activity was adjusted for every probe by dilution to 500,000 cpm/μl.

#### In situ hybridization

*In situ* hybridization was performed essentially as described previously.<sup>32</sup> In brief, 3 μm paraffin sections were deparaffinized in xylene, hydrated with graded ethanol and boiled in a microwave oven for 10–12 min in 10 mM citrate buffer, pH 6.0. After additional 20 min at room temperature, the sections were dehydrated with graded ethanol and the <sup>35</sup>S labeled probes (2×10<sup>6</sup> cpm in 20 μl hybridization mixture<sup>31</sup> per slide) incubated overnight at 55°C in a humidified chamber. Sections were washed in Hellen-dahl chambers with SSC buffers containing 0.1% SDS and 10 mM DTT at 150 rpm at 55°C using a Bühler incubation shaker (Johanna Guo GmbH, Hechingen, Germany) for 10 min in 2×SSC,

10 min in 0.5×SSC, and 10 min in 0.2×SSC. Sections were then RNase A treated for 10 min to remove nonspecifically bound riboprobe. Subsequent wash was performed in 0.2×SSC as specified above. Sections were dehydrated and soaked into an autoradiographic emulsion (Ilford), exposed for 5–7 days if not otherwise stated and finally developed. Sections were counterstained with haematoxylin and eosin.

#### Immunoperoxidase staining

Immunohistochemistry was performed essentially as described previously.<sup>32</sup> Five micrometer paraffin sections were deparaffinized with xylene and hydrated through ethanol/water dilutions. Tissue pretreatment was performed with protease-K (5 μg/ml) digestion for 20 min. Sections were blocked for endogenous peroxidase activity by treatment with 1% hydrogen peroxide for 15 min. The sections were washed in 50 mM Tris 150 mM NaCl, pH 7.6, containing 0.5% Triton X-100 (TBS-T). Incubation with antibodies was done overnight at 4°C. Sheep polyclonal antibodies (pAb) against TIMP-1 and nonimmune goat IgG were used at a final concentration of 4.0 μg/ml. Two monoclonal antibodies (MAb) against TIMP-1,<sup>33</sup> NM4 (clone rITX6A, NeoMarkers, Fremont, CA) and CalB2 (clone 147-6D11, CalBiochem, Oncogene Res. Products, Cambridge, MA), and a MAb against trinitrophenyl (TNP)<sup>34</sup> were all incubated at 1.0 μg/ml (all 3 MABs are IgG1). CalB2 MAB recognizes both free TIMP-1 and TIMP-1 in complex with MMPs.<sup>33</sup> NM4 MAB only recognizes free TIMP-1.<sup>33</sup> According to the manufacturer's descriptions, both MABs are raised using recombinant human TIMP-1. The sheep polyclonal antibodies were raised by immunization with TIMP-1 purified from human dermal fibroblasts. The IgG was obtained by triple precipitation using ammonium-sulfate and characterized by immunodiffusion and rocket immunoelectrophoresis.<sup>35</sup> In addition, we have shown that the pAb recognize both free and MMP-complexed TIMP-1.<sup>36</sup> Furthermore, the specificity of the antibodies was analyzed by Western blotting analysis against recombinant human TIMP-1 expressed in NSO mouse myeloma cells. Here, the antibody preparation recognizes a band of approximately 28 kDa in accordance with the molecular weight of TIMP-1. To certify that the pAb recognize TIMP-1 in colon tumors, the antibodies were immobilized on a sepharose column. Total protein extracted from 3 colon adenocarcinomas was passed through the column 5 times and the bound and subsequently eluted protein analyzed in a Western blot using a TIMP-1 monoclonal antibody (MAC15). A single band of approximately 28 kDa was revealed in accordance with the molecular weight of TIMP-1 (results not shown). In immunohistochemistry, the sheep pAb were detected with biotinylated rabbit-anti-goat IgG, which cross-react with sheep IgG (1:100, code E466, DakoCytomation) followed by horseradish peroxidase in complex with streptavidin (code K377, DakoCytomation). The MABs were detected with the Envision-mouse reagent (EnVision reagent, K4003, DakoCytomation), followed by tyramine amplification, using biotinyl tyramine substrate as specified by the manufacturer (Nen, Boston, MA). Sections were developed with NovaRed substrate as specified by the manufacturer (Vector Laboratories, Burlingame, CA) for 15 min. Finally, sections were counterstained in Mayers haematoxylin, dehydrated in ethanol and mounted.

#### Combined in situ hybridization and immunohistochemistry

Double labeling by combining *in situ* hybridization and immunohistochemistry on paraffin sections has been described previously.<sup>32</sup> In brief, using MAB against α-sm-actin (clone 1A4) diluted 1:1000, against cytokeratin (clone AE1/AE3) diluted 1:1000, or against CD68 (clone PGM1) diluted 1:200, sections were incubated for 2 hours at room temperature and then detected with anti-mouse-IgG/horse radish peroxidase-conjugated polymers (Envision-mouse reagent, DakoCytomation, Glostrup, Denmark). Sections were developed with diaminobenzidine (DAB) for 7–10 min, and immediately dehydrated for *in situ* hybridization, which was performed as described above using the antisense probes of f104. Sections were counterstained with haematoxylin.

## Results

### Analysis of TIMP-1 probes and antibodies for *in situ* hybridization and immunohistochemistry

Histopathological diagnosis of prospectively collected specimens from 18 colorectal lesions revealed 16 colorectal adenocarcinomas, 1 villous adenoma and 1 malignant lymphoma. <sup>35</sup>S-labeled antisense and sense RNA probes were generated by *in vitro* transcription from 2 nonoverlapping DNA sequences of the human TIMP-1 cDNA and tested by *in situ* hybridization on adjacent sections from 5 of the colorectal adenocarcinomas. The 2 antisense probes showed an identical hybridization pattern in all the 5 cases, located in the stromal compartment surrounding the invading cancer cells, while no specific signal was seen with the 2 sense probes (Fig. 1). To test whether the TIMP-1 mRNA was accompanied by TIMP-1 protein expression, immunohistochemistry was performed on 8 of the adenocarcinomas (including the 5 mentioned above) and the malignant lymphoma using sheep anti-human TIMP-1 polyclonal antibodies on sections adjacent to TIMP-1 *in situ* hybridized sections. The TIMP-1 mRNA and immunoreactivity was observed in the same cells in all of the 9 cases tested (Fig. 2A), including the malignant lymphoma. The anti-TIMP-1 polyclonal antibodies did not react with other cell populations in all of 8 adenocarcinomas and the malignant lymphoma apart from some normal and malignant epithelial cells that were weakly stained on the luminal apical surface. Two MABs against TIMP-1 (CalB2 and NM4) required strong signal amplification but showed a staining pattern similar to that of the polyclonal antibody preparation (Fig. 2B), with the only exception that neither of the 2 MABs stained the luminal apical surface of the normal and malignant epithelium. No signal was obtained with nonimmune goat serum or a MAB (of same subclass as CalB2 and NM4) directed against the synthetic hapten trinitrophenyl (TNP).

### TIMP-1 mRNA expression patterns in colon cancer

Expression of TIMP-1 mRNA was then analyzed in the remaining 9 colorectal lesions by *in situ* hybridization. TIMP-1 mRNA

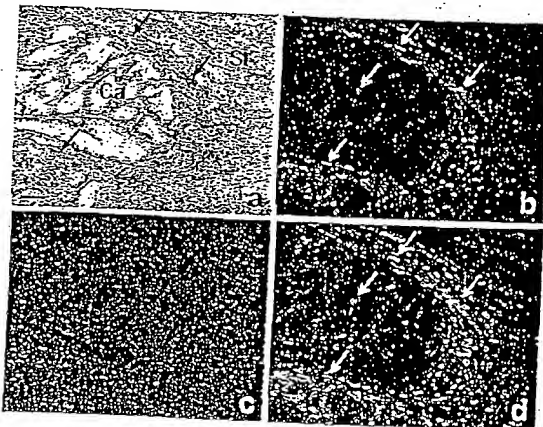


FIGURE 1 - *In situ* hybridization with 2 nonoverlapping TIMP-1 specific probes in human colon cancer. Three adjacent sections from a colon adenocarcinoma were incubated with 2 nonoverlapping <sup>35</sup>S-labeled antisense probes for TIMP-1 mRNA (H106 (a,b) and H104 (d)) and a corresponding TIMP-1 sense probe, H106 (c). The *in situ* hybridization signal is identified as black silver grains and demonstrated in brightfield (a) and as a white pattern in darkfield illumination (b-d). The 2 antisense probes show the same hybridization pattern and the hybridization signal is seen in the same cells (arrows in a, b and d), whereas no specific signal is seen with the sense probe (c). Note that the TIMP-1 mRNA signal is located in the tumor stroma (indicated by St) surrounding the invasive cancer cells (Ca) that are devoid of TIMP-1 *in situ* hybridization signal. (a)-(d): Bar = 100  $\mu$ m.

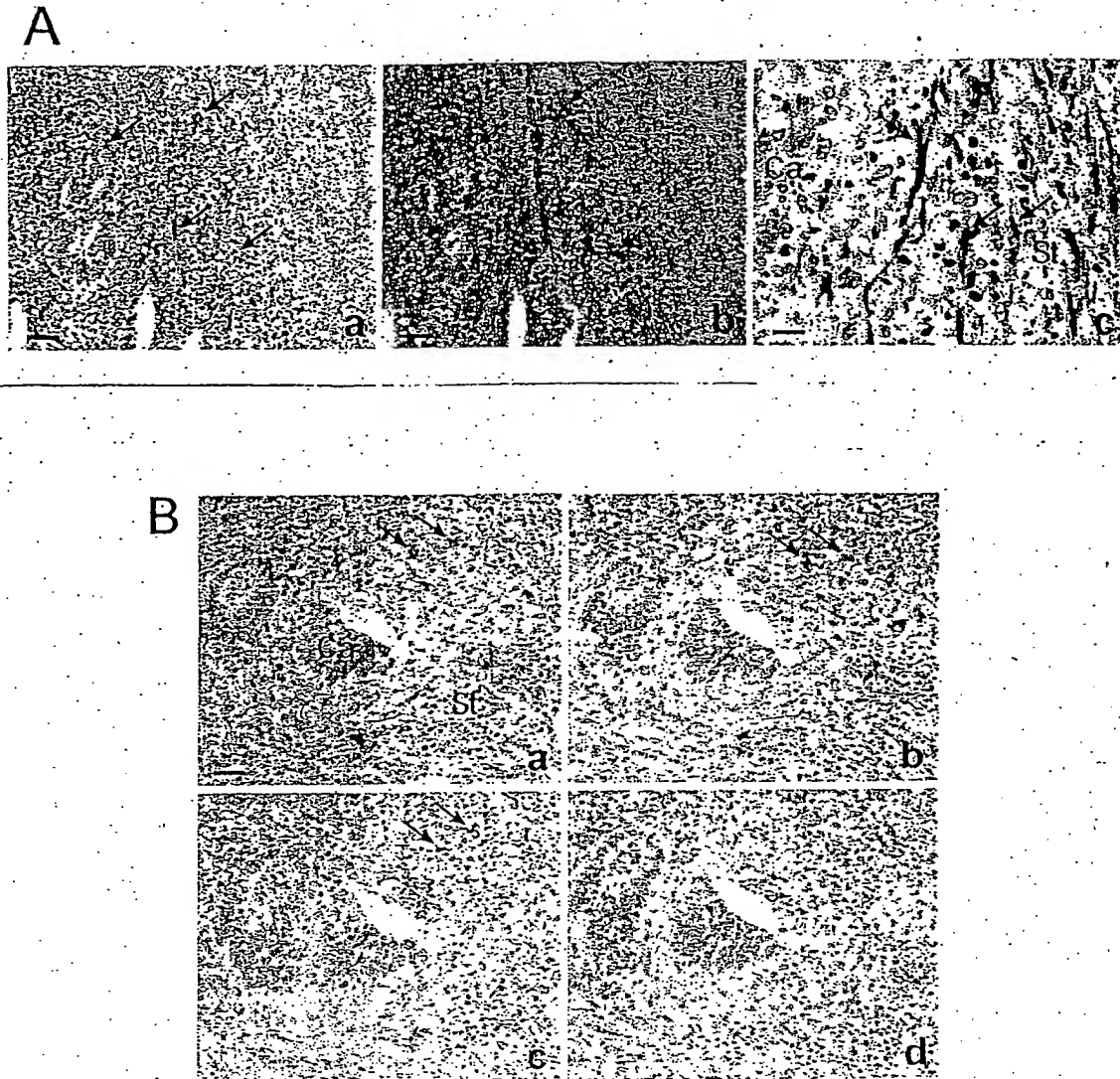
expression was in all the cases of colon adenocarcinoma (including those mentioned above) highly expressed in stromal fibroblast-like cells located at the invasive front (Fig. 3a,d). TIMP-1 mRNA signal was also observed in fibroblast-like cells located in the tumor stroma towards the colonic lumen in 8 of 10 cases where this tissue structure was present (data not shown). No or little TIMP-1 mRNA was detected in the central part of the carcinomas. In 5 of the 16 colorectal adenocarcinomas, we observed TIMP-1 mRNA signal in some fibroblast-like cells located around the muscle layer of some arteries located in the submucosa distant from the cancer area. The normal colonic mucosa, including the lamina propria that was present in all samples tested, was generally negative (Fig. 3b,e). Only a relatively weak TIMP-1 mRNA signal was detected in stromal fibroblast-like cells surrounding one or a very few normal crypts (Fig. 3c,f) in 3 out of 6 cases tested with extended exposure time (10 days vs. usually 5 days). In the villous adenoma, we saw only a few TIMP-1 mRNA positive cells associated with focal inflammation (data not shown). In the malignant lymphoma of the colon, TIMP-1 mRNA expressing fibroblast-like cells were, different from the adenocarcinomas, located in a diffuse pattern throughout the whole tumor. No TIMP-1 mRNA signal was observed in any of the 18 cases in the cancer cells, smooth muscle cells or vascular cells.

### Characterization of TIMP-1 mRNA expressing cells

To test whether the TIMP-1 mRNA positive fibroblast-like cells could be (myo)fibroblasts and/or macrophages, sections from 4 colorectal adenocarcinomas and the malignant lymphoma were first immunohistochemically stained with antibodies directed against  $\alpha$ -sm-actin [for detection of myofibroblast/smooth muscle cells (SMC)] or CD68 (for detection of macrophages) and subsequently incubated with a TIMP-1 mRNA antisense probe. In normal colon tissue,  $\alpha$ -sm-actin is expressed by vascular smooth muscle cells, smooth muscle cells of lamina muscularis mucosae and tunica muscularis as well as pericryptal myofibroblasts.<sup>37</sup> In colon tumors,  $\alpha$ -sm-actin is expressed by tumor-associated fibroblast-like cells located throughout the tumor stroma, which are defined as myofibroblasts. No TIMP-1 mRNA was detected in any  $\alpha$ -sm-actin positive smooth muscle cells, including those of the vessels, the lamina muscularis mucosae and the tunica muscularis. In addition, no TIMP-1 mRNA was detected in the  $\alpha$ -sm-actin positive pericryptal myofibroblasts of the lamina propria in any of the 5 lesions. TIMP-1 mRNA signal was in contrast seen in  $\alpha$ -sm-actin-positive tumor associated myofibroblasts located at the invasive front of the colon cancers. In 3 of the adenocarcinomas, more than 80% of TIMP-1 mRNA-positive cells located close to the invading cancer cells were  $\alpha$ -sm-actin-positive (Fig. 4). TIMP-1 mRNA positive fibroblast-like cells located more distant from the invasive cancer cells, towards the submucosa, expressed little or no  $\alpha$ -sm-actin. In 1 adenocarcinoma and in the malignant lymphoma approximately 50% of the TIMP-1 mRNA expressing cells constitute a subpopulation of tumor-associated myofibroblasts located at the invasive front of the tumor. No TIMP-1 mRNA signal could be identified in any of the CD68-positive cells (Fig. 4).

### Expression of TIMP-1 and MMP-2 and 9 in colon cancer

MMP-2 and MMP-9 are 2 type IV collagenases expressed in the invasive cancer tissue of colorectal adenocarcinomas. Thus MMP-2 has been reported to be expressed by fibroblast-like cells in the cancer stroma,<sup>12,31</sup> and MMP-9 by macrophages at the leading edge of the invasive cancer.<sup>33</sup> To directly compare the expression patterns of MMP-2 and MMP-9 with that of TIMP-1, adjacent sections from 5 colorectal adenocarcinomas were hybridized with probes for TIMP-1, MMP-2 and MMP-9 mRNAs. We found that the expression of TIMP-1 mRNA was localized characteristically at the invasive front of the growing tumor, whereas the expression of MMP-2 mRNA was most intense in the central areas, showing decreased expression towards the invasive front (Fig. 5a). MMP-9 mRNA expressing cells were found at the



**FIGURE 2**—*In situ* hybridization and immunohistochemistry for TIMP-1 in human colon cancer. (A) Two adjacent sections from a human colon adenocarcinoma were incubated with polyclonal antibodies against TIMP-1 (a,c) and a TIMP-1 mRNA antisense probe (b). The TIMP-1 immunoreactivity (red-brown color, arrows in a and c) and the TIMP-1 mRNA (silver grains, arrows in b) are identified in the same cells (arrows in a,b). Immunoperoxidase staining with the TIMP-1 pAb reveals the TIMP-1-positive cells as fibroblast-like cells (arrows in c) located in the stroma (St). No TIMP-1 immunoreactivity is seen in cancer cells (Ca). a,b: bars = 50  $\mu$ m; c: bars = 13  $\mu$ m. (B) Four consecutive adjacent sections were incubated with CalB2 MAB anti-TIMP-1 (a), NM4 MAB (b), sheep anti-TIMP-1 pAb (c) or mouse anti-TNP (d). The 3 MABs were detected with Envision reagent followed by TS amplification and the sheep pAb with biotinylated rabbit anti-goat followed by HRP-conjugated streptavidin (see Material and methods). The 3 TIMP-1 antibodies react with the same cells (arrows). No immunoreactivity is seen when the sections are incubated with anti-TNP.

invasive front like those expressing TIMP-1 mRNA but with a distinctly different distribution. Foci with high expression of TIMP-1 mRNA were not accompanied with increased expression of MMP-9 mRNA and vice-versa (Fig. 5b). Thus, TIMP-1 mRNA expression is not coregulated with MMP-2 or MMP-9 mRNA expression.

#### *TIMP-1 in adenomas and Dukes' stage A carcinomas*

TIMP-1 antigen can readily be measured in blood and we have previously reported that levels of TIMP-1 in blood are significantly

elevated in colorectal cancer patients compared to healthy donors and that high plasma TIMP-1 levels are associated with short survival of colorectal cancer patients.<sup>21,56</sup> TIMP-1 has therefore been suggested to be a novel marker for detection of early stage colorectal cancer and for prognostic stratification of colorectal cancer patients.<sup>21,39</sup> These findings, together with the characteristic expression pattern of TIMP-1 at the invasive front of virtually all the colon cancers and the absence or minute TIMP-1 expression in normal and benign colon mucosa, prompted the evaluation of TIMP-1 expression as a marker for early invasive colon cancer.



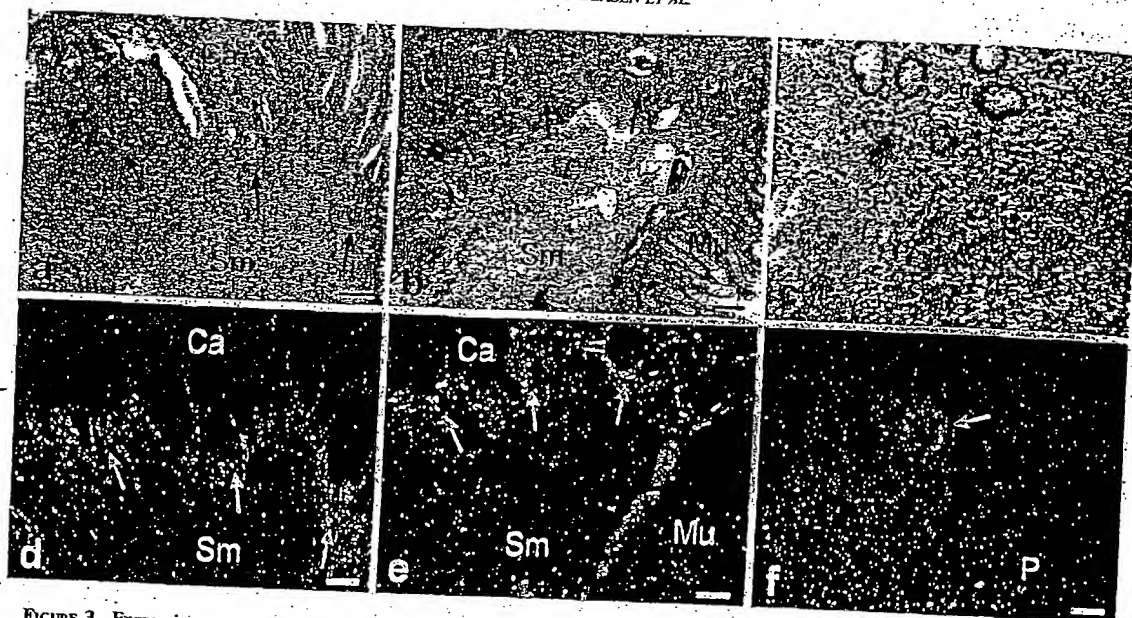


FIGURE 3—Expression patterns of TIMP-1 mRNA in human colon cancer. Sections were incubated with an antisense probe for TIMP-1 mRNA. The TIMP-1 mRNA signal is demonstrated in brightfield (a–c) and darkfield (d–f). The TIMP-1 mRNA is highly expressed in a subpopulation of stromal fibroblast-like cells at the invasive front of the cancer (a,d), whereas a low hybridization signal is seen within central parts of the cancer area (Ca) and in the submucosa (Sm). The normal mucosa (indicated by Mu in b and e) is negative. A low *in situ* hybridization signal was detected in a few stromal cells surrounding a few normal-looking glands (c,f), see also text. Exposure time: a,b,d,e, 5 days; d,f, 10 days. (a,b,d,f) bars = 200  $\mu$ m; (c,f) bars = 25  $\mu$ m.

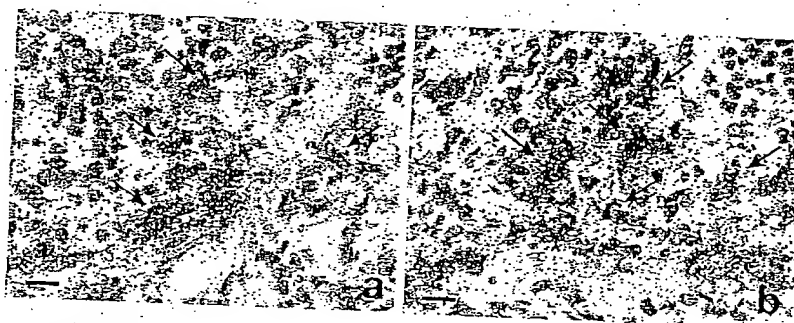


FIGURE 4—Double labeling for TIMP-1 mRNA and  $\alpha$ -sm-actin or CD68 in human colon cancer. Sections were first processed for immunohistochemistry incubating antibodies against  $\alpha$ -sm-actin (a) or CD68 (b) and subsequently by *in situ* hybridization using the f104 TIMP-1 specific antisense probe. TIMP-1 mRNA signal is colocalized with  $\alpha$ -sm-actin immunoreactivity in fibroblast-like cells that are considered as myofibroblasts (black arrows in a), whereas no CD68 immunoreactivity is seen in the TIMP-1 mRNA positive cells (black arrows in b) in an area with several CD68-positive macrophages (red arrows in b). Note that some of the  $\alpha$ -sm-actin-positive myofibroblasts have little or no TIMP-1 mRNA signal (red arrows in a). Bars = 13  $\mu$ m.

Therefore, we compared TIMP-1 mRNA expression in an additional 6 colorectal adenomatous polyps with the expression in an additional 8 Dukes' stage A colorectal adenocarcinomas. TIMP-1 mRNA signal was detected in 2 of the 6 adenomas, whereas all 8 Dukes' stage A carcinomas showed TIMP-1 mRNA signal at the invasive front (Fig. 6). In 1 positive adenoma (pedunculated type), TIMP-1 mRNA expression was confined to a single focus in fibroblast-like cells associated with focal stromal inflammation (Fig. 6). However, histological analysis of additional sections from this sample clearly revealed disruption of the dysplastic epithelium in the same area. In the other TIMP-1 mRNA positive adenoma (sessile type), a few TIMP-1 mRNA expressing fibroblasts were located around small arteries not directly associated with the tumor area (data not shown).

Taken together, all of the 9 Dukes' stage A carcinomas analyzed showed TIMP-1 mRNA expression in myofibroblasts located at the invasive front of the tumors, whereas expression was detected in only 3 of 7 adenomas, and in these was seen in fibroblast-like cells associated with focal inflammation at the epithelial-stromal interface in 2 of the cases and with arteries in the submucosa in 1 case.

#### Discussion

Our study was undertaken to clarify the expression and cellular localization of the MMP inhibitor TIMP-1 in human colon adenocarcinomas. Our studies were founded on the use of 2 specific antisense RNA probes derived from 2 nonoverlapping TIMP-1



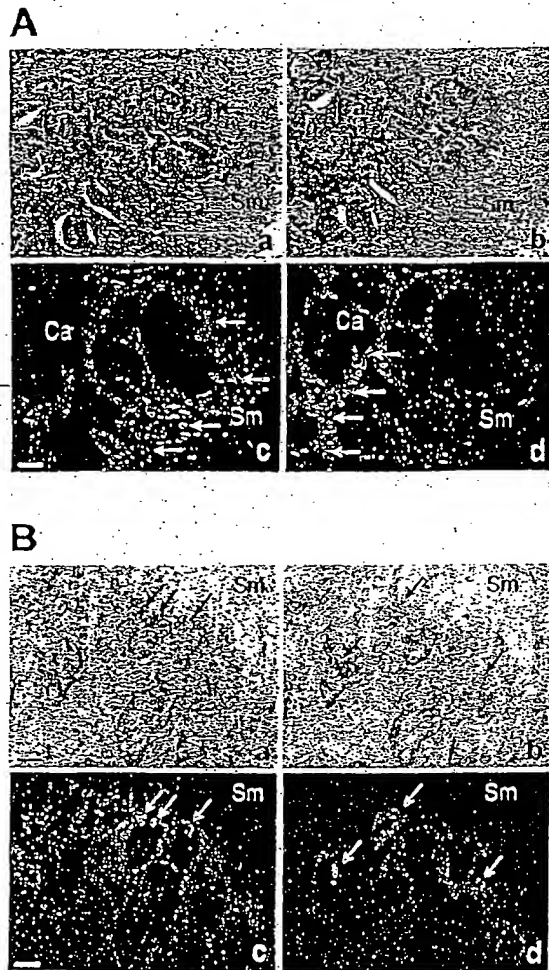


FIGURE 5—*In situ* hybridization for TIMP-1, MMP-2 and MMP-9 in human colon cancer. (A) Adjacent sections were incubated with probes specific for TIMP-1 mRNA (a,c) and MMP-2 mRNA (b,d), respectively, and is shown in brightfield (a,b) and darkfield illumination (c,d). The TIMP-1 mRNA signal increases towards the submucosa (Sm) whereas the MMP-2 mRNA signal decreases and is most intense in the central areas (Ca). (B) Adjacent sections were incubated with probes specific for TIMP-1 mRNA (a,c) and MMP-9 mRNA (b,d), respectively, and is here shown in brightfield (a,b) and darkfield illumination (c,d). Both the TIMP-1 mRNA signal and the MMP-9 mRNA signal are most intense at the invasive front towards the submucosa (Sm), but their expression patterns are quite different, with MMP-9 showing the most restricted expression. Bars = 100  $\mu$ m.

cDNA fragments and specific pAb and MAb against human TIMP-1. The TIMP-1 mRNA signal in all colorectal adenocarcinomas investigated was seen in fibroblast-like cells located in the tumor periphery. An identical hybridization pattern was observed with the 2 antisense TIMP-1 probes and application of complementary sense probes on neighboring tissue sections as negative controls did not result in any hybridization signal; therefore, we conclude that the hybridization signal generated with the antisense probes represents the genuine TIMP-1 mRNA. TIMP-1 immunoreactivity was also distinctly located in fibroblast-like stromal cells in the tumor periphery, and these cells were identified to be the same cells as the TIMP-1 mRNA expressing cells. A preparation

of sheep pAb against human TIMP-1<sup>35</sup> and 2 well-characterized MABs stained the very same cells in the tumor stroma. Weak staining of the apical surface of some normal and malignant epithelial cells was observed with the pAb in some of the samples. No staining was obtained when the anti-TIMP-1 antibodies were substituted with nonimmune goat serum or anti-TNP MAB incubated at the same concentrations. These immunohistochemical findings strongly suggest that the TIMP-1 antigen detected in the fibroblast-like cells represents the genuine TIMP-1 protein.

In our study, we found TIMP-1 mRNA expression in stromal fibroblast-like cells located in the tumor periphery in all colorectal adenocarcinomas tested, whereas no expression was detected in the cancer cells in any of the cases tested. This finding is in agreement with studies by Zeng and colleagues<sup>12,20</sup> but is partly in disagreement with findings by Newell and colleagues.<sup>27</sup> In addition to TIMP-1 mRNA signal in fibroblast-like cells in the tumor periphery, Newell and colleagues<sup>27</sup> detected a weak TIMP-1 mRNA signal in both benign and malignant epithelial cells.<sup>27</sup> This observation was, however, based on the use of probes from a single TIMP-1 cDNA subclone and no additional controls to verify the expression pattern. The difference between our results and those of Newell and colleagues may be explained by methodological differences, since the procedure employed by Newell and colleagues was in several steps different from the one used in the present study, e.g., Newell and colleagues used <sup>3</sup>H-labeled probes, whereas we used <sup>35</sup>S-labeled probes. It is in this context noteworthy that in order to look for a low expression level of TIMP-1 mRNA, we performed *in situ* hybridization experiments with prolonged exposure time (10 days vs. usually 5 days) with both our TIMP-1 antisense probes and both TIMP-1 sense probes, but with this challenge we did not detect any TIMP-1 mRNA in any epithelial cells. It cannot be excluded though that the TIMP-1 mRNA is expressed in epithelial cells below the detection limit of our *in situ* hybridization procedure.

An interesting observation in our study was the characteristic intense TIMP-1 mRNA and protein expression in the tumor periphery of all colon adenocarcinomas, while little or no expression was seen in the center of the carcinomas. Only in the colorectal lymphoma did we find TIMP-1 mRNA and protein expression in fibroblast-like cells located throughout the tumor tissue. The TIMP-1 expression pattern in the colon adenocarcinomas is in contrast to the expression pattern reported by Hewitt and colleagues,<sup>19</sup> who found that the TIMP-1 staining in most colorectal adenocarcinomas was equally intense in fibroblasts throughout the tumors and that some of the cases even showed decreased TIMP-1 signal intensity towards the tumor periphery. This difference may be explained by possible cross-reactivity of the polyclonal antibodies employed by Hewitt and colleagues or that Hewitt and colleagues employed cryostat sections, while we analyzed paraffin sections.

The TIMP-1 expressing cells had a fibroblast-like morphology and using combined *in situ* hybridization for TIMP-1 mRNA and immunohistochemistry for  $\alpha$ -sm-actin, we found that many, generally more than 50%, of the TIMP-1 mRNA positive cells coexpressed  $\alpha$ -sm-actin. According to the cellular morphology of the TIMP-1 expressing cells and their localization in the invasive front, we could conclude that the cells were myofibroblasts and not smooth muscle cells.

The myofibroblast is a cell type present in the normal colon mucosa, that originally was described as a pericryptal fibroblast<sup>37,40</sup> and later was identified with antibodies against  $\alpha$ -sm-actin.<sup>37</sup> In the lamina propria, the myofibroblasts form a continuous cell layer just below the intestinal epithelium. The pericryptal myofibroblasts are phenotypically different from the neighboring quiescent interstitial fibroblasts that do not express markers of smooth muscle cells.<sup>41</sup> During early steps of colonic tumorigenesis the number of myofibroblasts is significantly increased.<sup>41</sup> The TIMP-1 expressing myofibroblasts may be generated after activation of the pericryptal myofibroblasts and/or the quiescent inter-

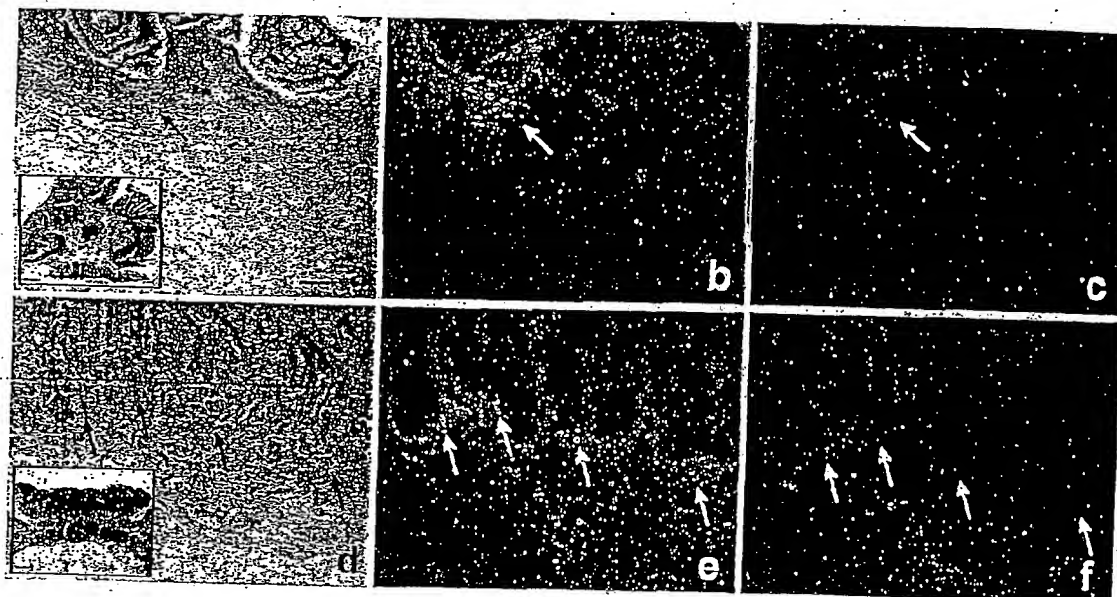


FIGURE 6—*In situ* hybridization for TIMP-1 in colon adenoma and Dukes' stage A adenocarcinoma. Sections from a moderate dysplastic sessile adenoma of the colon (a–c) and a Dukes' A colon carcinoma (d–f) were hybridized with TIMP-1 antisense (a, b and d, e) and sense (c, f) probes, shown in brightfield (a and d) and darkfield illumination (b, c and e, f). Inserts in a and d show virtually the whole tissue section with the arrow indicating the magnified area. TIMP-1 mRNA expression is confined to a single focus in the adenoma (arrows in a, b), whereas the TIMP-1 mRNA is detected along the invasive front in the Dukes' A colon carcinoma (arrows in d, e). (a)–(f): bar = 200  $\mu$ m. Bar in inserts = 1,250  $\mu$ m.

stitial fibroblasts. Adegboyega and colleagues<sup>41</sup> hypothesized that the tumor-associated myofibroblasts originate from the quiescent interstitial fibroblasts of the lamina propria, rather than from pericryptal myofibroblast or smooth muscle cells, which may help to explain why we found some of the TIMP-1 mRNA expressing fibroblast-like cells  $\alpha$ -sm-actin-positive and some  $\alpha$ -sm-actin-negative.

Several MMPs including MMP-2, MMP-11 and MMP-14 are expressed by fibroblast-like cells in human colon cancer,<sup>27,31,42–44</sup> some of which may indeed be myofibroblasts. The role of the (myo-) fibroblasts in colon cancer progression is not known. Since TIMP-1 in human colon cancer appears only to be expressed by fibroblast-like cells most of which are myofibroblasts, and since high TIMP-1 levels measured in blood or tumor extracts from colon cancer patients are strongly associated with a poor prognosis,<sup>20,21,27,28</sup> it could be argued that the TIMP-1 expressing myofibroblasts play a tumor-promoting role. Immunohistochemical localization studies of proteins involved in the activation and regulation of the efficient serine protease plasminogen, including urokinase plasminogen activator (uPA) and its specific inhibitor PAI-1 show that both are mainly expressed by myofibroblasts in human breast cancer.<sup>45,46</sup> High levels of uPA and PAI-1 are strongly correlated with poor prognosis in breast cancer,<sup>47,48</sup> supporting the assumption that the myofibroblast express a promoting role in cancer invasion. We recently reported that the predominant PAI-1 expressing cell in human colorectal cancer also is the myofibroblast,<sup>49</sup> and earlier studies indicated that elevated levels of PAI-1 in colon cancer patients are associated with poor prognosis.<sup>28</sup> Together these findings indicate that myofibroblasts are strongly contributing to the expression of proteins involved in the regulation of extracellular matrix degrading proteases that facilitate cancer invasion and metastasis.

A particularly interesting finding of the present study was the absence of TIMP-1 mRNA in 4 of 7 adenomas, whereas in all of 9 Dukes' stage A carcinomas the TIMP-1 mRNA was expressed in fibroblast-like cells along the invasive front. In the 2 benign lesions, in which the TIMP-1 mRNA was seen in the adenoma area, the

TIMP-1 mRNA positive cells were confined to a single focus with locally increased inflammation related to the dysplastic epithelium. Evident disruption of the dysplastic epithelium was observed in the adenoma with most intense TIMP-1 mRNA signal. Intestinal inflammation may be caused by disruption of the mucous epithelium that leads to focal leakage of mucinous colon material into the lamina propria. Increased intestinal permeability is a common deficiency in Crohn's disease and interestingly TIMP-1 mRNA was found in the intestinal granulation tissue of Crohn's disease<sup>50</sup> and is expressed by myofibroblasts isolated from Crohn's disease.<sup>51</sup> Induction of TIMP-1 in myofibroblasts in a benign or preinvasive tumor may also be a response to locally increased MMP activity or a response to the presence of a specific MMP in the local microenvironment. MMP-2 and MMP-9 mRNA expression, however, did not appear to be coregulated with TIMP-1 mRNA expression in the colorectal adenocarcinomas. Specific MMPs may indeed be involved in the transition of noninvasive to invasive disease; in studies of preinvasive lesions (ductal carcinomas *in situ*) of the human breast we recently reported that MMP-13 is specifically expressed in myofibroblasts associated with microinvasive events.<sup>32</sup> Future studies may clarify whether TIMP-1 expression in colorectal adenomas is correlated with expression of specific MMPs, cytokines and/or growth factors, such as TGF- $\beta$ 1 and TGF- $\beta$ 251, and whether TIMP-1 can be used as a histopathological marker for malignancy in colorectal tumors.

#### Acknowledgements

We thank Dr. G. Murphy for the TIMP-1 cDNA, the polyclonal antibodies against TIMP-1, MAC15 and recombinant TIMP-1. We are grateful to P.G. Knudsen and C. Lönborg for their excellent technical assistance, and to L.H. Engelholm for designing the primers. We appreciate the critical comments to the article by Dr. G. Kellermann.

## References

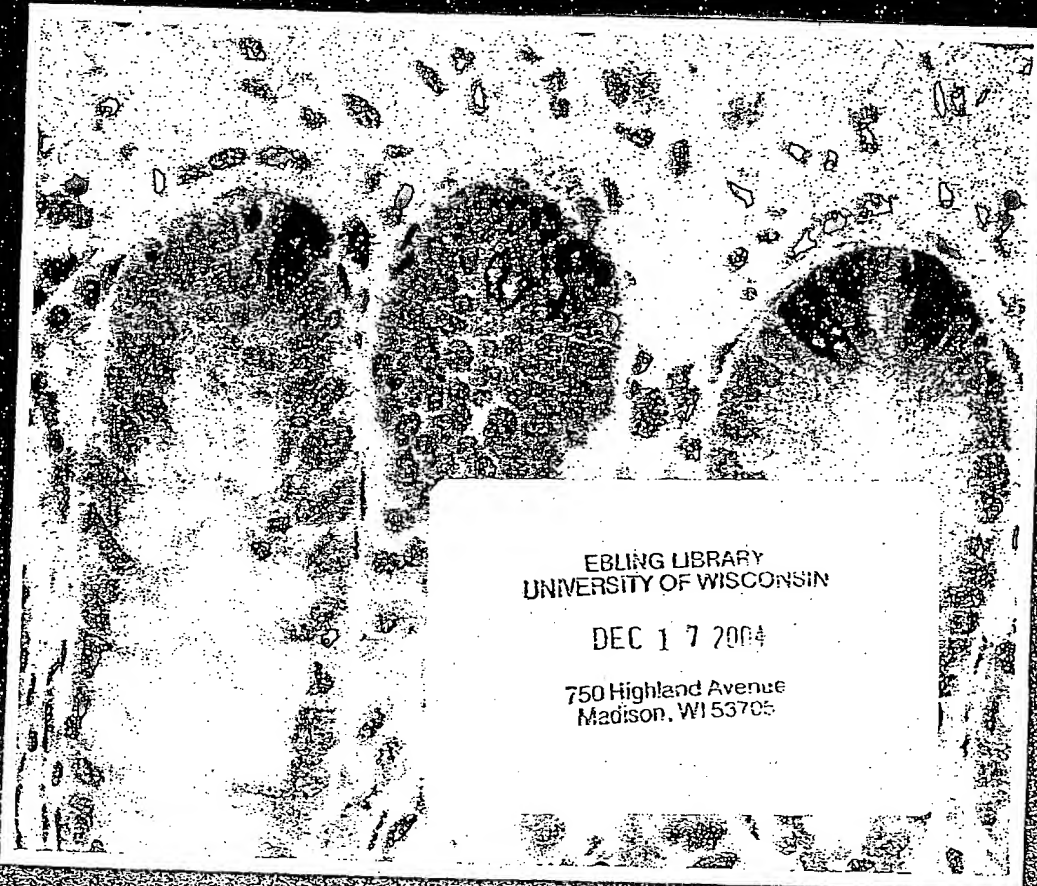
- Liotta LA, Stetler-Stevenson WG. Cancer metastasis and angiogenesis: an imbalance of positive and negative regulation. *Cell* 1991;64:327-36.
- Stetler-Stevenson WG, Yu AE. Proteases in invasion: matrix metalloproteinases. *Semin Cancer Biol* 2001;11:143-52.
- Murphy G, Kokkili S, Carne AF. Dissociation of tissue inhibitor of metalloproteinases (TIMP) from enzyme complexes yields fully active inhibitor. *Biochem J* 1989;261:1031-4.
- Welgus HG, Stricklin GP. Human skin fibroblast collagenase inhibitor. Comparative studies in human connective tissues, serum, and amniotic fluid. *J Biol Chem* 1983;258:12259-64.
- Khokha R, Waterhouse P. The role of tissue inhibitor of metalloproteinase-1 in specific aspects of cancer progression and reproduction. *J Neurooncol* 1999;18:123-7.
- Docherty AJ, Lyons A, Smith BJ, Wright EM, Stephens PE, Harris TJ, Murphy G, Reynolds JJ. Sequence of human tissue inhibitor of metalloproteinases and its identity to erythroid-potentiating activity. *Nature* 1985;318:66-9.
- Guedez L, Courtemanche L, Stetler-Stevenson M. Tissue inhibitor of metalloproteinase (TIMP)-1 induces differentiation and an antiapoptotic phenotype in germinal center B cells. *Blood* 1998;92:1342-9.
- Hewitt RE, Brown KE, Corcoran M, Stetler-Stevenson WG. Increased expression of tissue inhibitor of metalloproteinases type 1 (TIMP-1) in a more tumorigenic colon cancer cell line. *J Pathol* 2000;192:455-9.
- Guedez L, Stetler-Stevenson WG, Wolff L, Wang J, Fukushima P, Mansoor A, Stetler-Stevenson M. In vitro suppression of programmed cell death of B cells by tissue inhibitor of metalloproteinases-1. *J Clin Invest* 1998;102:2002-10.
- Li G, Fridman R, Kim HR. Tissue inhibitor of metalloproteinase-1 inhibits apoptosis of human breast epithelial cells. *Cancer Res* 1999;59:6267-75.
- Jiang Y, Goldberg ID, Shi YE. Complex roles of tissue inhibitors of metalloproteinases in cancer. *Oncogene* 2002;21:2245-52.
- Zeng ZS, Guillem JG. Distinct pattern of matrix metalloproteinase 9 and tissue inhibitor of metalloproteinase 1 mRNA expression in human colorectal cancer and liver metastases. *Br J Cancer* 1995;72:575-82.
- Gomez DE, Alonso DF, Yoshiji H, Thorgeirsson UP. Tissue inhibitors of metalloproteinases: structure, regulation and biological functions. *Eur J Cell Biol* 1997;74:111-22.
- Guillem JG, Levy MF, Hsieh LL, Johnson MD, LoGerfo P, Forde KA, Weinstein IB. Increased levels of p16, c-myc, and ornithine decarboxylase RNAs in human colon cancer. *Mol Carcinog* 1990;3:68-74.
- Lu XQ, Levy M, Weinstein IB, Santella RM. Immunological quantitation of levels of tissue inhibitor of metalloproteinase-1 in human colon cancer. *Cancer Res* 1991;51:6231-5.
- Mimori K, Mori M, Shiraiishi T, Fujie T, Baba K, Haraguchi M, Abe R, Ueo H, Akiyoshi T. Clinical significance of tissue inhibitor of metalloproteinase expression in gastric carcinoma. *Br J Cancer* 1997;76:531-6.
- Fong KM. TIMP1 and adverse prognosis in non-small cell lung cancer. *Clinical Cancer Research* 1996;2:1369-72.
- Ree AH, Flores VA, Berg JP, MacLandsmo GM, Nesland JM, Fodstad O. High levels of messenger RNAs for tissue inhibitors of metalloproteinases (TIMP-1 and TIMP-2) in primary breast carcinomas are associated with development of distant metastases. *Clin Cancer Res* 1997;3:1623-8.
- Hewitt RE, Leach JH, Powe DG, Clark IM, Cawston TE, Turner DR. Distribution of collagenase and tissue inhibitor of metalloproteinases (TIMP) in colorectal tumours. *Int J Cancer* 1991;49:666-72.
- Zeng ZS, Cohen AM, Zhang ZF, Stetler-Stevenson W, Guillem JG. Elevated tissue inhibitor of metalloproteinase 1 RNA in colorectal cancer stroma correlates with lymph node and distant metastases. *Clin Cancer Res* 1995;1:899-906.
- Holten-Andersen MN, Stephens RW, Nielsen HJ, Murphy G, Christensen JJ, Stetler-Stevenson W, Brummer N. High preoperative plasma tissue inhibitor of metalloproteinase-1 levels are associated with short survival of patients with colorectal cancer. *Clin Cancer Res* 2000;6:4292-9.
- Oberg A, Hoythya M, Tavelin B, Stenling R, Lindmark G. Limited value of preoperative serum analyses of matrix metalloproteinases (MMP-2, MMP-9) and tissue inhibitors of matrix metalloproteinases (TIMP-1, TIMP-2) in colorectal cancer. *Anticancer Res* 2000;20:1085-91.
- Pellegrini P, Contasta I, Berghella AM, Gargano E, Mammarella C, Adorno D. Simultaneous measurement of soluble carcinoembryonic antigen and the tissue inhibitor of metalloproteinase TIMP1 serum levels for use as markers of pre-invasive to invasive colorectal cancer. *Cancer Immunol Immunother* 2000;49:388-94.
- Grondahl-Hansen J, Christensen JJ, Rosenquist C, Brummer N, Mouridsen HT, Dano K, Blichert-Toft M. High levels of urokinase-type plasminogen activator and its inhibitor PAI-1 in cytosolic extracts of breast carcinomas are associated with poor prognosis. *Cancer Res* 1993;53:2513-21.
- Nielsen HJ, Pappot H, Christensen JJ, Brummer N, Thodacins-Ussing O, Moesgaard F, Dano K, Grondahl-Hansen J. Association between plasma concentrations of plasminogen activator inhibitor-1 and survival in patients with colorectal cancer. *BMJ* 1998;316:829-30.
- Wind T, Hansen M, Jensen JK, Andreassen PA. The molecular basis for anti-proteolytic and non-proteolytic functions of plasminogen activator inhibitor type-1: roles of the reactive centre loop, the shutter region, the flexible joint region and the small serpin fragment. *Biol Chem* 2002;383:21-36.
- Newell KJ, Witty JP, Rodgers WH, Matrisian LM. Expression and localization of matrix-degrading metalloproteinases during colorectal tumorigenesis. *Mol Carcinog* 1994;10:199-206.
- Tomita T, Iwata K. Matrix metalloproteinases and tissue inhibitors of metalloproteinases in colonic adenomas-adenocarcinomas. *Dis Colon Rectum* 1996;39:1255-64.
- O'Shea M, Willenbrock F, Williamson RA, Cockeitt MI, Freedman RB, Reynolds JJ, Docherty AJ, Murphy G. Site-directed mutations that alter the inhibitory activity of the tissue inhibitor of metalloproteinases-1: importance of the N-terminal region between cysteine 3 and cysteine 13. *Biochemistry* 1992;31:10146-52.
- Engelholm LH, Nielsen BS, Netzel-Arnett S, Solberg H, Chen XD, Lopez Garcia JM, Lopez-Otin C, Young MF, Birkedal-Hansen H, Dano K, Lund LR, Behrendt N, et al. The urokinase plasminogen activator receptor-associated protein/endostatin is coexpressed with its interaction partners urokinase plasminogen activator receptor and matrix metalloproteinase-13 during osteogenesis. *Lab Invest* 2001;81:1403-14.
- Pyke C, Rallkier E, Huhtala P, Hurskainen T, Dano K, Tryggvason K. Localization of messenger RNA for Mr 72,000 and 92,000 type IV collagenases in human skin cancers by in situ hybridization. *Cancer Res* 1992;52:1336-41.
- Nielsen BS, Rank F, Lopez JM, Balbin M, Vizoso F, Lund LR, Dano K, Lopez-Otin C. Collagenase-3 expression in breast myofibroblasts as a molecular marker of transition of ductal carcinoma in situ lesions to invasive ductal carcinomas. *Cancer Res* 2001;61:7091-100.
- Holten-Andersen MN, Brummer N, Maimonis P, Jensen V, Murphy G, Piironen T. Characterization of monoclonal antibodies to tissue inhibitor of metalloproteinases-1. *Journal of clinical ligand assay* 2002;25:87-90.
- Shulman M, Wilde CD, Kohler G. A better cell line for making hybridomas secreting specific antibodies. *Nature* 1978;276:269-70.
- Hembry RM, Murphy G, Reynolds JJ. Immunolocalization of tissue inhibitor of metalloproteinases (TIMP) in human cells. Characterization and use of a specific antiserum. *J Cell Sci* 1985;73:105-19.
- Holten-Andersen MN, Murphy G, Nielsen HJ, Pedersen AN, Christensen JJ, Hoyer-Hansen G, Brummer N, Stephens RW. Quantitation of TIMP-1 in plasma of healthy blood donors and patients with advanced cancer. *Br J Cancer* 1999;80:495-503.
- Sappino AP, Dietrich PY, Skalli O, Widgeon S, Gabbiani G. Colonic pericyptal fibroblasts. Differentiation pattern in embryogenesis and phenotypic modulation in epithelial proliferative lesions. *Virchows Arch A Pathol Anat Histopathol* 1989;415:551-7.
- Nielsen BS, Timshel S, Kjeldsen L, Sehested M, Pyke C, Borregaard N, Dano K. 92 kDa type IV collagenase (MMP-9) is expressed in neutrophils and macrophages but not in malignant epithelial cells in human colon cancer. *Int J Cancer* 1996;65:57-62.
- Holten-Andersen MN, Christensen JJ, Nielsen HJ, Stephens RW, Jensen V, Nielsen OH, Sorensen S, Overgaard J, Lilja H, Harris A, Murphy G, Brummer N. Total levels of tissue inhibitor of metalloproteinases 1 in plasma yield high diagnostic sensitivity and specificity in patients with colon cancer. *Clin Cancer Res* 2002;8:156-64.
- Kaye GI, Lane N, Pascal RR. Colonic pericyptal fibroblast sheath: replication, migration, and cytodifferentiation of a mesenchymal cell system in adult tissue. II. Fine structural aspects of normal rabbit and human colon. *Gastroenterology* 1968;54:852-65.
- Adegboyega PA, Milfin RC, DiMari JF, Saada JJ, Powell DW. Immunohistochemical study of myofibroblasts in normal colonic mucosa, hyperplastic polyps, and adenomatous colorectal polyps. *Arch Pathol Lab Med* 2002;126:829-36.
- Poulsen R, Pignatelli M, Stetler-Stevenson WG, Liotta LA, Wright PA, Jeffery RE, Longcroft JM, Rogers L, Stamp GW. Stromal expression of 72 kDa type IV collagenase (MMP-2) and TIMP-2 mRNAs in colorectal neoplasia. *Am J Pathol* 1992;141:389-96.
- Rouyer N, Wolf C, Chenard MP, Rio MC, Chambou P, Bellocq JP, Basset P. Stromelysin-3 gene expression in human cancer: an overview. *Invasion Metastasis* 1994;14:269-75.

44. Okada A, Bellocq JP, Ronyer N, Chevard MP, Rio MC, Chambon P, Basset P. Membrane-type matrix metalloproteinase (MT-MMP) gene is expressed in stromal cells of human colon, breast, and head and neck carcinomas. *Proc Natl Acad Sci U S A* 1995;92:2730-4.
45. Nielsen BS, Sehested M, Dunn S, Rank F, Timshel S, Rygaard J, Johnsen M, Dano K. Urokinase plasminogen activator is localized in stromal cells in ductal breast cancer. *Lab Invest* 2001;81:1485-501.
46. Offeren BV, Nielsen BS, Høyer-Hansen G, Rank F, Hamilton-Dutoit S, Overgaard J, Andreassen PA. The myofibroblast is the predominant plasminogen activator inhibitor-1 expressing cell type in human breast carcinomas. *Am J Pathol* 2003;163:1887-99.
47. Duffy MJ, Reilly D, O'Sullivan C, O'Higgins N, Fennelly JJ, Andreassen P. Urokinase-plasminogen activator, a new and independent prognostic marker in breast cancer. *Cancer Res* 1990;50:6827-9.
48. Grondahl HJ, Christensen JJ, Rosenquist C, Brunner N, Mouridsen HT, Dano K, Blichert TM. High levels of urokinase-type plasminogen activator and its inhibitor PAI-1 in cytosolic extracts of breast carcinomas are associated with poor prognosis. *Cancer Res* 1993;53:2513-21.
49. Illemann M, Hansen U, Nielsen HJ, Andreassen PA, Høyer-Hansen G, Lund LR, Dano K, Nielsen BS. Leading edge myofibroblasts in human colon cancer express PAI-1. *Am J Clin Pathol* 2004;122:256-65.
50. Saarialho-Kere UK, Vaalamo M, Prolakksinen P, Airola K, Parks WC, Karjalainen-Lindsberg ML. Enhanced expression of matrilysin, collagenase, and stromelysin-1 in gastrointestinal ulcers. *Am J Pathol* 1996;148:519-526.
51. McKaig BC, McWilliams D, Watson SA, Mahida YR. Expression and regulation of tissue inhibitor of metalloproteinase-1 and matrix metalloproteinases by intestinal myofibroblasts in inflammatory bowel disease. *Am J Pathol* 2003;162:1355-1360.

Q  
7IN84115

INTERNATIONAL

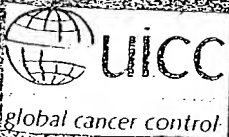
# Journal of Cancer



EBLING LIBRARY  
UNIVERSITY OF WISCONSIN

DEC 17 2004

750 Highland Avenue  
Madison, WI 53705



Publication of the International Union Against Cancer  
Published for the UICC by Wiley-Blackwell

Articles published online in Wiley InterScience

25 August 2004 - 8 October 2004

[www.internationaljournalofcancer.com](http://www.internationaljournalofcancer.com)

ISSN: 0020-7179



global cancer control

## International Union Against Cancer

### COUNCIL

J. SEFFRIN, President (USA)  
L.J. DENIS, Treasurer (Belgium)  
J. BAIT, Chair, Finance Committee (USA)

G. BRIEN (Australia)  
R.C. BURTON (Australia)  
M. DAUBE (Australia)  
K.A. DINGSHAW (India)  
L. ELOVAINIO (Finland)  
M.K. GOSPODAROWICZ (Canada)  
R.T. HUDSON (Ireland)

T. KITAGAWA (Japan)  
A. PUNDALIK KURKURE (India)  
R.E. LENHARD (USA)  
A. LLOMBART-BOSCH (Italy)

M. LUWIA (Indonesia)  
C. MALLINSON (UK)  
L.H. MARCHESI (Brazil)  
H.F. MICKELSON (USA)  
K. NILSSON (Sweden)  
S. OMAR (Egypt)  
T. PHILIP (France)  
E. ROBINSON, Past President (Israel)  
Y. SALOOJEE (South Africa)  
H. SANCHE-GARNIER (France)

R.J. SCHWEITZER (USA)  
O. SOREIDE (Norway)  
K. TAJIMA (Japan)  
W. WEBER (Switzerland)  
S. WILKINSON (UK)  
D. ZACKS (USA)  
D. ZARIDZE (Russia)  
Y. HUI ZHANG (China)  
M. ZIV (Israel)  
H. ZUR HAUSEN (Germany)

UICC, URL: <http://www.uicc.org>

UICC, E-mail: [info@uicc.org](mailto:info@uicc.org)

The International Union Against Cancer (UICC) is devoted exclusively to all aspects of the world-wide fight against cancer. Its objectives are to advance scientific and medical knowledge in research, diagnosis, treatment and prevention of cancer, and to promote all other aspects of the campaign against cancer throughout the world. Particular emphasis is placed on professional and public education.

Founded in 1933, the UICC is a non-governmental, independent association of more than 290 member organizations in over 80 countries. Members are voluntary cancer leagues and societies, cancer research and/or treatment centers, and in some countries ministries of health.

The UICC is non-profit, non-political, and non-sectarian. Its headquarters are in Geneva, Switzerland. It creates and carries out programs around the world in collaboration with hundreds of volunteer experts. Supported by membership dues, national subscriptions, grants and donations, its annual budget is about US \$4 million.

The UICC is governed by its members which meet in General Assembly every 4 years. Its elected Council and Executive Committee are responsible for Program structure and implementation.

The UICC organizes an International Cancer Congress every 4 years as well as annual symposia, workshops, and training courses. It publishes the *International Journal of Cancer* (30 issues per year), *UICC News* (quarterly), the *International Calendar of Meetings on Cancer* (bi-annually), and a number of technical reports, textbooks and manuals.

### SUBSCRIPTION INFORMATION

© 2005 Wiley-Liss, Inc., a Wiley Company. All rights reserved. No part of this publication may be reproduced in any form or by any means, except as permitted under section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the publisher, or authorization through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, telephone: (978) 750-8400, fax: (978) 750-4470. Requests to the publisher for permission should be addressed to the Permissions Department, c/o John Wiley & Sons, Inc., 111 River St., Hoboken, NJ 07030. Fax: (201) 748-6008; Tel.: (201) 748-6011; <http://www.wiley.com/go/permissions>.

*International Journal of Cancer* (Print ISSN 0020-7136; Online ISSN 1097-0215) is published 30 times a year, semi-monthly with extra issues in January, March, May, July, September, and November, by Wiley-Liss, Inc., through Wiley Subscription Services, Inc., a Wiley Company. Send subscription inquiries in care of John Wiley & Sons, Inc., Attn: Journals Admin Dept UK, 111 River St., Hoboken, NJ 07030, (201) 748-6645.

Advertising inquiries should be addressed to Advertising Department, c/o John Wiley & Sons, Inc., 111 River St., Hoboken, NJ 07030. Telephone: (201) 748-6921.

Offprint sales and inquiries should be directed to the Customer Service Department, in care of John Wiley & Sons, Inc., 111 River St., Hoboken, NJ 07030. Telephone: (201) 748-8776.

Subscription price: Volumes 113-117, 2005 (30 issues). Print only: \$2,905 worldwide. Electronic only: \$2,905 worldwide. A combination price of \$3,195 worldwide includes the subscription in both electronic and print formats. A special personal rate is available to individuals for \$295 worldwide. All subscriptions containing a print element, shipped outside the U.S., will be sent by air. Payment must be made in U.S. dollars drawn on U.S. bank.

Periodicals postage paid at Hoboken, NJ and at additional mailing offices. Postmaster: send address changes to INTERNATIONAL JOURNAL OF CANCER, Subscription Distribution, c/o John Wiley & Sons, Inc., 111 River St., Hoboken, NJ 07030. Change of address: Please forward to the subscriptions address listed above 6 weeks prior to move; enclose present mailing label with change of address. Claims for missing issues: Claims for undelivered copies will be accepted only after the following issue has been received. Please enclose a mailing label or cite your subscriber reference number. Missing copies will be supplied when losses have been sustained in transit and where reserve stock permits. Send claims in care of John Wiley & Sons, Inc., Attn: Journals Admin Dept UK, 111 River St., Hoboken, NJ 07030.

Indexed by: EMBASE/Excerpta Medica • Current Contents/Life Sciences • Science Citation Index • Scisearch • BIOSIS Data Base • Index Medicus • Cambridge Scientific Abstracts • Chemical Abstracts • Reference Update • Smoking and Health Database.

This journal is printed on acid-free paper.

### INTERNATIONAL JOURNAL OF CANCER

2005 Wiley-Liss, Inc.

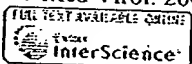
The *International Journal of Cancer* is published for the International Union Against Cancer by Wiley-Liss, Inc., a division of John Wiley & Sons, Inc. Five volumes are issued annually, each consisting of six numbers.

Abstracting, and other journals may reprint the summaries of articles without requesting authorization. Authors alone are responsible for views expressed in signed articles. The mention of specific companies or of certain manufacturers' products does not imply that they are endorsed or recommended by the International Union Against Cancer.

References: When quoting from the *International Journal of Cancer*, please use the official abbreviation:

*Int. J. Cancer*





**Tissue plasminogen activator induced by dengue virus infection of human endothelial cells.**

Huang YH, Lei HY, Liu HS, Lin YS, Chen SH, Liu CC, Yeh TM.

Department of Microbiology and Immunology, College of Medicine, National Cheng Kung University, Tainan, Taiwan, ROC.

Dengue hemorrhagic fever and dengue shock syndrome (DHF/DSS) are severe complications of dengue virus (DV) infection. However, the pathogenesis of hemorrhage induced by dengue virus infection is poorly understood. Since endothelial cells play a pivotal role in the regulation of hemostasis, we studied the effect of DV infection on the production of tissue plasminogen activator (tPA) and plasminogen activator inhibitor 1 (PAI-1) in vitro using both primary isolated endothelial cells; human umbilical cord veins cells, and a human microvascular endothelial cell line. DV infection significantly induced the secretion of tPA but not PAI-1 of human endothelial cells. In addition, tPA mRNA of endothelial cells was induced by DV as demonstrated by RT-PCR. Antibody against IL-6 but not control antibody inhibited DV-induced tPA production of endothelial cells. Furthermore, a good correlation between sera levels of IL-6 and tPA was found in DHF but not DF patients. These results suggest that IL-6 can regulate DV-induced tPA production of endothelial cells, which may play important roles in the pathogenic development of DHF/DSS. Copyright 2003 Wiley-Liss, Inc.

PMID: 12794725 [PubMed - indexed for MEDLINE]

**Neu oncogene expression in ovarian tumors: a quantitative study.**

Huettner PC, Carney WP, Naber SP, DeLellis RA, Membrino W, Wolfe HJ.

Department of Pathology, Tufts University School of Medicine, Massachusetts.

We studied neu mRNA expression by slot blot analysis and protein product expression by capture ELISA and immunohistochemistry in 57 primary and metastatic ovarian neoplasms, two paraovarian leiomyosarcomas, and eight normal ovaries. Some 61% of ovarian tumors but none of the paraovarian neoplasms or normal ovaries overexpressed neu mRNA. A total of 96% of the ovarian tumors that overexpressed neu were of epithelial type. Epithelial ovarian tumors had significantly higher amounts of the neu oncogene product as determined by capture ELISA than either germ cell and stromal tumors or normal ovaries ( $p$  less than 0.025). Different subtypes of ovarian carcinomas had significantly different amounts of neu oncogene product as measured by capture ELISA; endometrioid tumors had the highest, and poorly differentiated carcinomas not otherwise specified had the lowest ( $p$  less than 0.025). ELISA values, mRNA overexpression, and immunohistochemical staining intensity did not correlate with stage at diagnosis or architectural or nuclear grade in ovarian tumors. We conclude that capture ELISA is a simple, effective way to measure the neu oncogene protein product and that there is a good correlation between ELISA levels and immunohistochemical staining intensity. However, ELISA values did not correlate with stage or histologic prognostic factors in ovarian neoplasms.

PMID: 1353878 [PubMed - indexed for MEDLINE]



**Real-time quantitative RT-PCR of cyclin D1 mRNA in mantle cell lymphoma: comparison with FISH and immunohistochemistry.**

Hui P, Howe JG, Crouch J, Nimmakayalu M, Qumsiyeh MB, Tallini G, Flynn SD, Smith BR.

Department of Laboratory Medicine, Yale University School of Medicine, 333 Cedar Street, P.O. Box 208035, New Haven, CT 06520-8035, USA.

Presence of the balanced translocation t(11;14)(q13;q32) and the consequent overexpression of cyclin D1 found in mantle cell lymphoma (MCL) has been shown to be of important diagnostic value. Although many molecular and immunohistochemical approaches have been applied to analyze cyclin D1 status, correlative studies to compare different methods for the diagnosis of MCL are lacking. In this study, we examined 39 archived paraffin specimens from patients diagnosed with a variety of lymphoproliferative diseases including nine cases meeting morphologic and immunophenotypic criteria for MCL by: (1) real-time quantitative RT-PCR to evaluate cyclin D1 mRNA expression; (2) dual fluorescence in situ hybridization (FISH) to evaluate the t(11;14) translocation in interphase nuclei; and (3) tissue array immunohistochemistry to evaluate the cyclin D1 protein level. Among the nine cases of possible MCL, seven cases showed overexpression of cyclin D1 mRNA (cyclin D1 positive MCL) and two cases showed no cyclin D1 mRNA increase (cyclin D1 negative "MCL-like"). In six of seven cyclin D1 positive cases, the t(11;14) translocation was demonstrated by FISH analysis; in one case FISH was unsuccessful. Six of the seven cyclin D1 mRNA overexpressing cases showed increased cyclin D1 protein on tissue array immunohistochemistry; one was technically suboptimal. Among the two cyclin D1 negative MCL-like cases, FISH confirmed the absence of the t(11;14) translocation in both cases. All other lymphoproliferative diseases studied were found to have low or no cyclin D1 mRNA expression and were easily distinguishable from the cyclin D1 overexpressing MCLs by all three techniques. In addition, to confirming the need to assess cyclin D1 status, as well as, morphology and immunophenotyping to establish the diagnosis of MCL, this study demonstrates good correlation and comparability between measure of cyclin D1 mRNA, the 11;14 translocation and cyclin D1 protein.

Publication Types:

- Evaluation Studies

PMID: 12952233 [PubMed - indexed for MEDLINE]

**Elevation of topoisomerase I messenger RNA, protein, and catalytic activity in human tumors: demonstration of tumor-type specificity and implications for cancer chemotherapy.**

Husain I, Mohler JL, Seigler HF, Besterman JM.

Department of Cell Biology, Glaxo Inc. Research Institute, Research Triangle Park, North Carolina 27709.

Topoisomerase I has been identified as an intracellular target of camptothecin, a plant alkaloid with anticancer activity. Various lines of evidence suggest that the sensitivity of cells to this drug is directly related to the topoisomerase I content. In humans, the levels of topoisomerase I have been shown to be elevated in colorectal tumors, compared to normal colon mucosa. The aim of our study was to determine whether (a) topoisomerase I levels are elevated in other solid tumors, (b) the elevated enzyme is catalytically active in these tumors, and (c) the increase in topoisomerase I levels in colorectal tumors is a result of increased transcription or translation. Topoisomerase I levels were quantitated in crude extracts from colorectal, prostate, and kidney tumors and their matched normal counterparts by Western blotting and by direct determination of catalytic activity, and mRNA levels were determined by Northern blotting. By Western blotting, colorectal tumors showed 5-35-fold increases in topoisomerase I levels, compared to their normal colon mucosa. In the case of prostate tumors, the increase was 2-10-fold, compared with benign hyperplastic prostate tissue from the same patients. However, no difference was observed in topoisomerase I levels in kidney tumors, compared to their normal counterparts. The catalytic activity of topoisomerase I was determined by a quantitative <sup>32</sup>P-transfer assay in crude homogenates, without isolating nuclei. Colorectal and prostate tumors exhibited 11-40- and 4-26-fold increases, respectively, in catalytic activity. However, kidney tumors did not show any alteration in catalytic activity, compared to their normal matched samples. Thus, for all three tumor types there was a good correlation between enzyme levels and catalytic activity. Finally, colorectal tumors were analyzed for steady state mRNA levels. A 2-33-fold increase in mRNA levels was found in colorectal tumors, compared to normal colon mucosa. These results suggest that alterations in topoisomerase I expression in humans are tumor type specific and that the increase in topoisomerase I levels results from either increased transcription of the topoisomerase I gene or increased mRNA stability.

PMID: 8275492 [PubMed - indexed for MEDLINE]

**High-level mRNA quantification of proliferation marker pKi-67 is correlated with favorable prognosis in colorectal carcinoma.**

Ihmann T, Liu J, Schwabe W, Hausler P, Behnke D, Bruch HP, Broll R, Windhovel U, Duchrow M.

St. Elisabeth Klinik, Klinik für Anesthesiologie, Schmerztherapie und Intensivmedizin, Saarlouis, Germany.

**PURPOSE:** The present study retrospectively examines the expression of pKi-67 mRNA and protein in colorectal carcinoma and their correlation to the outcome of patients. **METHODS:** Immunohistochemistry and quantitative RT-PCR were used to analyze the expression of pKi-67 in 43 archival specimens of patients with curatively resected primary colorectal carcinoma, who were not treated with neo-adjuvant therapy. **RESULTS:** We determined a median pKi-67 (MIB-1) labeling index of 31.3% (range 10.3-66.4%), and a mean mRNA level of 0.1769 (DeltaC(T): range 0.01-0.69); indices and levels did not correlate. High pKi-67 mRNA DeltaC(T) values were associated with a significantly favorable prognosis, while pKi-67 labeling indices were not correlated to prognostic outcome. A multivariate analysis of clinical and biological factors indicated that tumor stage (UICC) and pKi-67 mRNA expression level were independent prognostic factors. **CONCLUSION:** Quantitatively determined pKi-67 mRNA can be a good and new prognostic indicator for primary resected colorectal carcinoma.

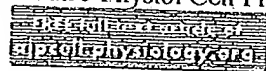
**Publication Types:**

- [Evaluation Studies](#)

PMID: 15449182 [PubMed - indexed for MEDLINE]

187: Am J Physiol Cell Physiol. 2001 Oct;281(4):C1396-402.

Related Articles, Links



**Modulation of glucagon receptor expression and response in transfected human embryonic kidney cells.**

Ikegami T, Cypess AM, Bouscarel B.

Department of Medicine, George Washington University Medical Center, Washington, District of Columbia 20037, USA.

The modulation of glucagon receptor (GR) expression and biological response was investigated in human embryonic kidney cell (HEK-293) clones permanently expressing the GR with different densities. The GR mRNA expression level in these clones was upregulated by cellular cAMP accumulation and presented a good correlation with both the protein expression level and the maximum number of glucagon binding sites. However, the determination of glucagon-induced cAMP accumulation in these cell lines revealed that the enhancement of receptor expression did not lead to a proportional increase in cAMP formation. Under these conditions, the maximum cAMP production induced by NaF and forskolin was not significantly different among selected clones, regardless of the receptor expression level. High receptor-expressing clones showed the greatest susceptibility for agonist-induced desensitization compared with clones with lower GR expression levels. The results of the present study suggest that the GR can recruit non-GR-specific desensitization mechanism(s). Furthermore, the partial inhibition or alteration of the overall cAMP synthesis pathway at the receptor level may be a necessary adaptive step for a cell in response to a massive increase in membrane receptor expression level.

PMID: 11546678 [PubMed - indexed for MEDLINE]

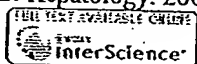
**Developmental regulation of acidic fibroblast growth factor (aFGF) expression in bovine retina.**

Jacquemin E, Jonet L, Oliver L, Bugra K, Laurent M, Courtois Y, Jeanny JC.

Unite de Recherches Gerontologiques, U. 118 INSERM, Paris, France.

Acidic fibroblast growth factor (aFGF) is a signalling molecule implicated in a wide variety of biological processes such as cell growth, differentiation and survival. It has been purified from bovine retina. The present study was carried out to detect which cells in the bovine retina expressed aFGF at the different stages of embryonic and post-natal development. The specific aFGF mRNA and protein were detected by in situ hybridization employing riboprobes and immunocytochemistry using affinity purified polyclonal human recombinant aFGF antibodies respectively. No signal was detected by either technique until 4-5 months and then there was progressive expression of aFGF with terminal morphogenesis of the retina. By 8-9 months of embryonic development, nuclei of the 3 neuronal layers (ganglion cell layer, inner and outer nuclear layers) were all uniformly and intensely labeled. A slight labeling of the pigmented epithelium of the retina was also visible throughout development and maturation. These results showed a good correlation between message and protein expression in these cell types. In contrast, glial cells in the nerve fiber layer and vascular endothelial cells displayed a nuclear immunostaining for the protein in the absence of message. These data suggest that aFGF plays a role in the late steps of retinal differentiation by autocrine and paracrine mechanisms.

PMID: 7507349 [PubMed - indexed for MEDLINE]



**The p21(Cip1) protein, a cyclin inhibitor, regulates the levels and the intracellular localization of CDC25A in mice regenerating livers.**

**Jaime M, Pujol MJ, Serratos J, Pantoja C, Canela N, Casanovas O, Serrano M, Agell N, Bachs O.**

Department of Cell Biology and Pathology, Faculty of Medicine, Institut d'Investigacions Biomediques August Pi Sunyer (IDIBAPS), University of Barcelona, Barcelona, Spain.

Liver cells from p21(Cip1<sup>-/-</sup>) mice subjected to partial hepatectomy (PH) progress into DNA synthesis faster than those from wild-type mice. These cells also show a premature induction of cyclin E/cyclin-dependent kinase (CDK) 2 activity. We studied the mechanisms whereby cells lacking p21(Cip1) showed a premature induction of this activity. Whereas the levels of CDK2, cyclin E, and p27(Kip1) were similar in both wild-type and p21(Cip1<sup>-/-</sup>) mice, those of the activator CDC25A were much higher in p21(Cip1<sup>-/-</sup>) quiescent and regenerating livers than in wild-type animals. Moreover, p21(Cip1<sup>-/-</sup>) cells also showed a premature translocation of CDC25A from cytoplasm into the nucleus. The ectopic expression of p21(Cip1) into mice embryo fibroblasts from p21(Cip1<sup>-/-</sup>) mice decreased the levels of CDC25A and delayed its nuclear translocation. The levels of CDC25A messenger RNA in p21(Cip1<sup>-/-</sup>) cells were higher than in wild-type cells, suggesting that this increase might be responsible, at least in part, for the high levels of CDC25A protein in these cells. Thus, the results reported here indicate that p21(Cip1) regulates the levels and the intracellular localization of CDC25A. We also found a good correlation between CDC25A nuclear translocation and cyclin E/CDK2 activation. In conclusion, premature translocation of CDC25A to the nucleus might be involved in the advanced induction of cyclin E/CDK2 activity and DNA replication in cells from animals lacking p21(Cip1).

PMID: 11981756 [PubMed - indexed for MEDLINE]



### Alteration of frizzled expression in renal cell carcinoma.

Jaussens N, Audries L, Janicot M, Perera T, Bakker A.

Department of Biochemistry, University of Antwerp, Wilrijk, Belgium.  
njansse9@prdbe.jnj.com

To evaluate the involvement of frizzled receptors (Fzds) in oncogenesis, we investigated mRNA expression levels of several human Fzds in more than 30 different human tumor samples and their corresponding (matched) normal tissue samples, using real-time quantitative PCR. We observed that the mRNA level of Fzd5 was markedly increased in 8 of 11 renal carcinoma samples whilst Fzd8 mRNA was increased in 7 of 11 renal carcinoma samples. Western blot analysis of crude membrane fractions revealed that Fzd5 protein expression in the matched tumor/normal kidney samples correlated with the observed mRNA level. Wnt/beta-catenin signaling pathway activation was confirmed by the increased expression of a set of target genes. Using a kidney tumor tissue array, Fzd5 protein expression was investigated in a broader panel of kidney tumor samples. Fzd5 membrane staining was detected in 30% of clear cell carcinomas, and there was a strong correlation with nuclear cyclin D1 staining in the samples. Our data suggested that altered expression of certain members of the Fzd family, and their downstream targets, could provide alternative mechanisms leading to activation of the Wnt signaling pathway in renal carcinogenesis. Fzd family members may have a role as a biomarker.

PMID: 15557753 [PubMed - indexed for MEDLINE]

1. A3  
054  
v. 25  
no. 4  
2004  
Jul-Aug

1010-4283

Received on: 01-23  
Tumour biology: the journal  
of the International Society  
for Oncodevelopmental  
Biology and Medicine.

2004 | 4 | 04

July-August 2004  
(Released November 2004)  
ISSN 1010-4283  
25(4) 157-220 (2004)

The Journal of the International Society for  
Oncodevelopmental Biology and Medicine

# Tumor Biology

Tumor Markers, Tumor Targeting  
and Translational Cancer Research

## Research Articles

- 157 Tissue Microarray Analysis of Cyclin D1 Gene Amplification and Gain in Colorectal Carcinomas  
Toncheva, D.; Petrova, D.; Tzenova, V.; Dimova, I.; Yankova, R.; Yordanov, V.; Damjanov, D.; Todorov, T.; Zaharieva, B. (Sofia)
- 161 Alteration of Frizzled Expression in Renal Cell Carcinoma  
Janssens, N. (Wilrijk/Beerse); Andries, L. (Edgem); Janicot, M.; Perera, T.; Bakker, A. (Beerse)
- 172 Antisense and Dominant-Negative AKT2 cDNA Inhibits Glioma Cell Invasion  
Pu, P.; Kang, C.; Li, J. (Tianjin); Jiang, H. (Detroit, Mich.)
- 179 Production and Characterization of a New Antibody Specific for the Mutant EGF Receptor, EGFRvIII, in *Camelus bactrianus*  
Omidfar, K.; Rasati, M.J. (Tehran); Modjtahedi, H. (Guilford); Forouzandeh, M.; Taghikhani, M.; Bakhtiani, A.; Paknejad, M.; Kashanian, S. (Tehran)
- 188 Matrix Metalloproteinases 2 and 9 and Their Tissue Inhibitors in Low Malignant Potential Ovarian Tumors  
Määttä, M.; Santala, M.; Soini, Y.; Tolvensaari-Mattila, A.; Turpeenniemi-Hujanen, T. (Oulu)

- 193 Human Kallikrein 6 Degrades Extracellular Matrix Proteins and May Enhance the Metastatic Potential of Tumour Cells  
Ghosh, M.C.; Grass, L.; Soosaipillai, A. (Toronto); Sotiropoulou, G. (Patras); Diamandis, E.P. (Toronto)

## Mini Reviews

- 200 The *TP53* Tumor Suppressor Gene and Melanoma Tumorigenesis: Is There a Relationship?  
Hussein, M.R. (Assuit)
- 208 Strategies to Endow Cytotoxic T Lymphocytes or Natural Killer Cells with Antibody Activity against Carcinoembryonic Antigen  
Kuroki, M.; Kuroki, M.; Shibaguchi, H.; Badran, A.; Hachimine, K.; Zhang, J.; Kinugasa, T. (Fukuoka)

## Research Commentary

- 217 Up Close and Personal: Molecular Diagnostics in Oncology  
Rye, P.D. (Oslo); Nilsson, O. (Göteborg); Rittenhouse, H. (San Diego, Calif.); Stigbrand, T. (Umeå)

S. Karger  
Medical and Scientific  
Publishers  
Basel · Freiburg  
Paris · London  
New York · Bangalore  
Bangkok · Singapore  
Tokyo · Sydney

KARGER

## Online

Access to full text and tables of contents,  
including tentative ones for forthcoming issues:  
[www.karger.com/tbi\\_issues](http://www.karger.com/tbi_issues)



## Alteration of Frizzled Expression in Renal Cell Carcinoma

Nico Janssens<sup>a,c</sup> Luc Andries<sup>b</sup> Michel Janicot<sup>c</sup> Tim Perera<sup>c</sup>  
Annette Bakker<sup>c</sup>

<sup>a</sup>Department of Biochemistry, University of Antwerp, Wilrijk, <sup>b</sup>HistoGeneX, Edegem, and <sup>c</sup>Oncology Discovery Research, Johnson & Johnson Pharmaceutical Research and Development, Beerse, Belgium

### Key Words

$\beta$ -Catenin · Cyclin D1 · Frizzled receptor · Renal cell carcinoma · Wnt

### Abstract

To evaluate the involvement of frizzled receptors (Fzds) in oncogenesis, we investigated mRNA expression levels of several human Fzds in more than 30 different human tumor samples and their corresponding (matched) normal tissue samples, using real-time quantitative PCR. We observed that the mRNA level of Fzd5 was markedly increased in 8 of 11 renal carcinoma samples whilst Fzd8 mRNA was increased in 7 of 11 renal carcinoma samples. Western blot analysis of crude membrane fractions revealed that Fzd5 protein expression in the matched tumor/normal kidney samples correlated with the observed mRNA level. Wnt/ $\beta$ -catenin signaling pathway activation was confirmed by the increased expression of a set of target genes. Using a kidney tumor tissue array, Fzd5 protein expression was investigated in a broader panel of kidney tumor samples. Fzd5 membrane staining was detected in 30% of clear cell carcinomas, and there was a strong correlation with nuclear cyclin D1 staining in the samples. Our data suggested that altered expression of certain members of the

Fzd family, and their downstream targets, could provide alternative mechanisms leading to activation of the Wnt signaling pathway in renal carcinogenesis. Fzd family members may have a role as a biomarker.

Copyright © 2004 S. Karger AG, Basel

### Introduction

The Wnt signaling pathway is evolutionary conserved and controls many events during embryonic development. Members of the Wnt gene family of secreted glycoproteins are involved in embryonic induction, generation of cell polarity, cell proliferation and the determination of cell fate [1, 2]. Recently, it has become evident that the Wnt pathway is also deregulated in a range of tumors [3].

The Wnt signaling pathway is activated when Wnt proteins bind to a cell surface receptor complex consisting of a member of the frizzled receptor (Fzd) family and either low-density-lipoprotein receptor-related protein (LRP)5 or LRP6 [4, 5]. A detailed characterization of the Fzds and the immediate downstream events after Wnt binding has been hampered by the lack of pure biologically active Wnts.

Downstream of the receptor complex, three pathways may be initiated, depending on the composition of the

KARGER

Fax +41 61 306 12 34  
E-Mail karger@karger.ch  
www.karger.com

© 2004 S. Karger AG, Basel  
1010-2283/04/0254-0161\$21.00/0

Accessible online at:  
www.karger.com/tbi

Nico Janssens  
Oncology Discovery Research, JNJPRD  
Turnhoutseweg 30  
BE-2340 Beerse (Belgium)  
Tel. +32 14 605831. Fax +32 14 605403. E-Mail njanss@prdbe.jnj.com

ligand and receptor complex. The 'Wnt/ $\beta$ -catenin pathway', the 'Wnt/ $\text{Ca}^{2+}$  pathway' or the 'Wnt polarity pathway' [6]. The Wnt/ $\beta$ -catenin pathway has been linked to carcinogenesis. Genetic alterations in components of this pathway (adenomatous polyposis coli, APC, axin and  $\beta$ -catenin) can result in the accumulation of non-phosphorylated  $\beta$ -catenin [3, 7] and this can promote carcinogenesis. Conversely, neither the Wnt/ $\text{Ca}^{2+}$  pathway nor the Wnt polarity pathway involves the activation of  $\beta$ -catenin [for review, see ref. 1, 6].

Mutations in one of the three regulatory genes (APC,  $\beta$ -catenin and axin), overexpression of Wnts and Fzds or the expression of a constitutively active Fzd have been linked to Wnt/ $\beta$ -catenin pathway activation in various tumors [8, 9].

To evaluate the involvement of Fzds in oncogenesis, we investigated mRNA expression levels of several human Fzds (Fzd2, 3, 5, 6, 7, 8 and 9) in more than 30 different human tumor samples using real-time quantitative PCR. Each sample was compared with its corresponding (matched) normal tissue sample. The most striking observation was the dramatically increased Fzd5 and Fzd8 mRNA expression seen in the renal carcinoma samples. This was confirmed at the protein level using Western blotting. Kidney tumor tissue arrays confirmed Fzd5 membrane staining in 30% of clear cell carcinomas, with nuclear cyclin D1 showing a strong correlation with the Fzd5 membrane labeling. Fzd8 protein expression analysis was not performed due to the lack of suitable reagents. These data suggest that Fzd5 may have a role in renal cell carcinogenesis due to its frequent overexpression observed in these tumor samples. Potential future applications could include uses in tumor targeting or as a potential biomarker.

## Materials and Methods

### Tissue Samples

Frozen tumor tissue samples with corresponding normal tissue from the same patient were derived either from human biopsy or autopsy material (Department of Pathology, University of Antwerp, kindly provided by Prof. E. Van Marck). Tissue specimens were snap-frozen in liquid nitrogen and kept at  $-80^{\circ}\text{C}$  until use. Frozen sections of kidney tumor and normal tissue samples were stained with hematoxylin-eosin to support the pathologist's observations and to confirm the type of kidney tumor. Paraffin-embedded tissue slides of renal carcinoma, lung carcinoma, breast and colon carcinoma were obtained, after encryption, from the Department of Pathology (Middelheim Hospital, Antwerp, Belgium). The CLLI human kidney cancer (SuperBioChips Laboratories) tissue array used in this study contained 59 tissue samples consisting of 9 normal kidney tissues,

30 clear cell renal carcinoma samples and another 20 renal cell tumor types (chromophil, chromophobe, papillary type, collecting duct carcinoma and samples with mixed types).

### RNA Isolation and Reverse Transcription

Total RNA was extracted from tissue specimens using Ultraspec Reagent (Biotecx, USA) according to the manufacturer's instructions. All total RNA was routinely treated with DNase (DNA-free kit, Ambion, USA). 1  $\mu\text{g}$  of total RNA was used to synthesize cDNA using oligo-dT primers (Superscript; Invitrogen, Merelbeke, Belgium). Reverse transcription was performed at  $42^{\circ}\text{C}$  for 60 min, followed by  $70^{\circ}\text{C}$  for 10 min.

### Real-Time PCR

Real-time PCR was performed on either an ABI Prism 7700 or 7900 Sequence detection system (Perkin-Elmer Applied Biosystems, Foster City, Calif., USA) using the 5' nuclease assay (Taqman<sup>TM</sup>). Primer and probe sequences were designed using Primer Express (PE Applied Biosystems) and are shown in table 1. Quantitative values were obtained from the threshold cycle number (Ct) at which the increase in the signal associated with exponential growth of PCR products is detected using PE Biosystems analysis software, according to the manufacturer's instructions.

We have used the  $2^{-\Delta\Delta\text{Ct}}$  method to analyze the relative changes in gene expression of the different genes between tumor and corresponding normal tissue samples. We used the mitochondrial ATP synthase 6 (ATP6) as the endogenous RNA control [10; Janssens et al., in prep.], and each sample was normalized to its ATP6 content. The relative expression of the target gene was also normalized to the corresponding normal tissue sample (calibrator). Results, expressed as the amount of target sample relative to the ATP6 gene and the calibrator, were determined as follows:  $N = 2^{-\Delta\text{Ct}(\text{sample}) - \Delta\text{Ct}(\text{calibrator})}$ , where the  $\Delta\text{Ct}$  values of the sample and calibrator were determined by subtracting the average Ct value of the sample and the calibrator from the average Ct value of the ATP6 gene. Amplification was done essentially as described previously [10]. Briefly, 50  $\mu\text{l}$  of reaction mixture containing 1  $\mu\text{l}$  of cDNA template were amplified as follows: incubation at  $50^{\circ}\text{C}$  for 2 min; denaturation at  $95^{\circ}\text{C}$  for 10 min, and 50 cycles at  $95^{\circ}\text{C}$  for 15 s and  $60^{\circ}\text{C}$  for 1 min.

### Membrane Preparation, Gel Electrophoresis and Immunoblotting

Tissue samples were weighed, suspended at a 40 times dilution [ $= 40$  volumes/original wet weight of tissue (v/w)] in 50 mM Tris-HCl buffer, pH 7.4, and homogenized with an Ultra-Turrax homogenizer. After centrifugation for 10 min, 24,000 g at  $4^{\circ}\text{C}$ , the pellet was washed three times by resuspension in the Tris-HCl buffer followed by centrifugation. The final membrane pellets were stored at  $-80^{\circ}\text{C}$  in the Tris-HCl buffer at a concentration of 0.5–1 mg/ml. The Bradford protein assay (Pierce, Aalst, Belgium) was used for protein determination. Proteins (50  $\mu\text{g}$ ) were separated by 8% SDS-PAGE and transferred to nitrocellulose membranes. After primary and secondary antibody incubation, the antigen-antibody-peroxidase complex was detected by chemiluminescence (Pierce, Aalst, Belgium) according to the manufacturer's instructions.

### Immunohistochemistry

Immunohistochemistry was performed on 10- $\mu\text{m}$ -thick cryosections of unfixed tumor tissue and on 6- $\mu\text{m}$ -thick paraffin sections from renal tumor tissue fixed by formalin or by an alcohol-based fixative. Adjacent tissue blocks from renal tumors were processed with

Table 1. Real-time PCR primer and probe sequences

Target cDNA	Primer/probe sequences <sup>a</sup>	Fragment position <sup>b</sup>	Accession No. <sup>c</sup>
FZD2	(a) 5'-atcccgigcccgge-3' (b) 5'-gtattgacatgagatcggaagtc-3' (c) 5'-FAM-tacacgcccgcgcatgicgc-TAMRA-3'	1,548-1,613	AB017364
FZD3	(a) 5'-icacgctcagtgcatggg-3' (b) 5'-ttgtacacitcaatitlaticatcg-3' (c) 5'-FAM-catcccggaactctaacatcatccitit-TAMRA-3'	1,473-1547	AB039723
FZD5	(a) 5'-tgcacaggtcacttccggtt-3' (b) 5'-tcttcaagtcgcccgcg-3' (c) 5'-FAM-ctttcagtggtcgttgcctcc-TAMRA-3'	2,143-2,204	HSU43318
FZD6	(a) 5'-ctagcaccctcagggaagagaa-3' (b) 5'-ctcagagagatcggagatggal-3' (c) 5'-FAM-tgggtgaacatgcctcgcag-TAMRA-3'	2,094-2,170	AF072873
FZD7	(a) 5'-ccttggaaaggcataacttg-3' (b) 5'-aaccaacgggaaccicaga-3' (c) 5'-FAM-aagcaactttataggcaagcagtgcaa-TAMRA-3'	2,687-2,762	AB017365
FZD8	(a) 5'-tgggtcgggtccttctt-3' (b) 5'-cgctccatgctgataaggaag-3' (c) 5'-FAM-ccaccttcgcaecgtctcca-TAMRA-3'	853-919	AB043703
FZD9	(a) 5'-ccccgggagctacggac-3' (b) 5'-tagtcatgctgcaagaccacgg-3' (c) 5'-FAM-tggcagcactgccaataaggct-TAMRA-3'	1,696-1,763	HSU82169
ATP5b	(a) 5'-gggtgagggtgcttgggt-3' (b) 5'-gggcgcagtgattataggctt-3' (c) 5'-FAM-aagtggttagggcattttaaacttagagcg-TAMRA-3'	580-503	AF368271
c-myc	(a) 5'-accaccagcagcgccttga-3' (b) 5'-tccagcagaaggatccagact-3' (c) 5'-FAM-accctttgcaggagccttccct-TAMRA-3'	1,297-1,413	HSMYCI
Cyclin D1	(a) 5'-gaacttgccgcaatgac-3' (b) 5'-cgctcttggaatttggga-3' (c) 5'-FAM-cgcacgatttcaatgaacatt-TAMRA-3'	4,148-4,211	AF511593
PPAR $\delta$	(a) 5'-agcatctcaccggcaaa-3' (b) 5'-gtctcgatgctggatcaca-3' (c) 5'-FAM-ccagccacacgggcct-TAMRA-3'	932-990	NM-006238

<sup>a</sup> (a) = Sense primer; (b) = antisense primer; (c) = probe.

<sup>b</sup> Fragment positions are given according to the EMBL/GenBank accession No. of cloned sequence.

<sup>c</sup> EMBL/GenBank accession No. of cloned sequence.

formalin and with the alcohol-based fixative. Paraffin and cryosections were mounted on poly-L-lysine or 3-aminopropyltriethoxysilane-gelatin-coated slides. The 59 tissue samples on the CLI human renal cancer tissue array slides were all fixed with formalin and embedded in paraffin, and the sections were mounted on silane-coated slides (SuperBioChips Laboratories). In addition to renal carcinoma tissue, sections from 10 formalin-fixed paraffin-embedded lung carcinomas were stained for Fzd5. Colon and breast tumors

were used as positive controls for  $\beta$ -catenin and cyclin D1 immunostaining.

The following primary antibodies were used: Fzd5 (Upstate Biotechnology),  $\beta$ -catenin (Zymed), cyclin D1 (Zymed), E-cadherin (Novocastra) and cytokeratin 8 (Biogenex). Cryosections were fixed in 4% paraformaldehyde for 5 min; acetone for 5 min at  $-20^{\circ}\text{C}$  and 70% ethanol for 5 min. Endogenous peroxidase activity was quenched using 3%  $\text{H}_2\text{O}_2$ . Paraffin sections of formalin- and alcohol-

Table 2. Fzd mRNA expression in tumor samples

Sample <sup>a</sup>	Tissue	Tumor type	x-fold expression increase <sup>b</sup>						
			FZD2	FZD3	FZD5	FZD6	FZD7	FZD8	FZD9
133702	kidney	adenocarcinoma	0.17	0.5	3.72	2.13	0.06	1.32	-
137770	kidney	renal cell carcinoma	1.23	3.56	8.26	2.61	1.3	8.21	5.44
138844	kidney	renal cell carcinoma	0.31	0.11	6.84	1.18	0.23	3.18	2.8
137146	kidney	renal cell carcinoma	0.47	0.45	3.16	2.23	0.73	4.42	1.37
137564	kidney	renal cell carcinoma	3.43	2.95	9.6	1.57	7.64	3.52	-
133408	kidney	renal cell carcinoma	23.97	0.98	6.39	2.48	5.05	16.72	-
139188	kidney	renal cell carcinoma	3.7	0.56	0.66	1.33	0.37	4.41	1.53
135699	kidney	renal cell carcinoma	2.6	0.36	4.83	0.9	0.34	2.54	2.31
139064	kidney	renal cell carcinoma	5.16	1.82	1.25	2.36	1.8	2.19	2.85
134585	kidney	renal cell carcinoma	1.47	0.72	1.38	0.47	0.09	0.6	-
140279	kidney	renal cell carcinoma	7.33	18.17	3.93	6.05	6.41	4.65	7.91
137252	ovary	carcinosarcoma	0.4	3.39	0.49	0.44	0.92	0.54	0.53
138256	ovary	papillary carcinoma	0.7	5.17	1.22	2.19	0.4	0.11	-
146472	ovary	serous papillary carcinoma	0.39	3.29	2.56	1.34	3.94	0.67	0.59
145845	colon	adenocarcinoma	3.45	3.36	0.58	1.22	1.67	1.24	1.99
146145	colon	adenocarcinoma	5.46	6.74	4.42	6.57	3.36	6.08	0.15
146630	colon	adenocarcinoma	4.01	4.73	0.3	1.16	0.54	0.55	0.12
146633	colon	adenocarcinoma	1.87	1.07	1.62	1.47	0.51	2.07	-
147055	colon	adenocarcinoma	0.66	1.41	1.01	1.33	0.24	0.79	4.87
142253	lung	adenocarcinoma	0.67	5.17	0.43	1.87	0.33	0.59	0.41
143036	lung	adenocarcinoma	0.67	1.24	2.63	1.13	1.11	1.04	9.19
138938	lung	adenocarcinoma	0.93	0.88	1.07	1.3	1.73	0.38	0.76
133563	lung	adenocarcinoma	2.76	1.23	0.31	1.34	0.99	0.78	4.97
144387	lung	adenocarcinoma	12.85	0.43	0.49	0.25	2.73	0.54	-
137304	lung	acinary adenocarcinoma	0.54	9.15	1.65	1.25	0.63	3.41	2.64
144546	lung	epithelial carcinoma	0.1	0.67	0.21	8.11	0.27	0.97	1.32
137621	lung	epithelial carcinoma	1.52	2.19	0.37	2.26	0.52	0.45	1.06
145552	lung	epithelial carcinoma	1.09	-	0.16	0.91	1.47	0.13	0.5
143987	testis	embryonal carcinoma	0.24	0.6	19.43	1.33	23.12	2.43	0.33
137332	stomach	leiomyoma	19.32	3.91	0.45	1.53	18.98	32.82	16.7
139026	stroma	gastrointestinal carcinoma	66.1	3.51	0.03	6.67	8.44	2.28	20.82
136049	rectum	adenocarcinoma	1.71	0.52	0.4	0.54	0.29	0.88	0.28
140794	gall bladder	adenosquamous carcinoma	0.19	-	0.93	3.02	0.2	-	-

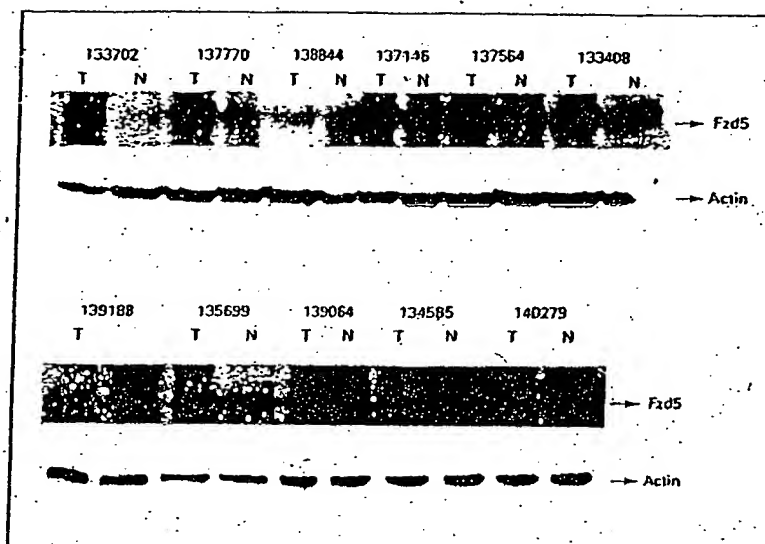
<sup>a</sup> Sample identification numbers were given by the pathologist.

<sup>b</sup> Results are expressed as x-fold increase of the gene in the tumor tissue sample compared to its matched normal tissue sample after normalizing both samples on the basis of their ATP5b content. A cutoff of 3-fold was used to define differential expression. Significant (> 3-fold) increases in the expression level of the Fzd receptor are shown in italics. - = Expression of the target gene undetectable in one or both samples (tumor and/or normal).

fixed tissue were processed with a trypsin-citrate-microwave pretreatment or with an EDTA-microwave pretreatment to unmask epitopes, respectively. Sections were then sequentially processed with primary antibodies, biotinylated secondary antibodies and streptavidin-biotin-peroxidase (Fzd5, E-cadherin and cytokeratin 8). For  $\beta$ -catenin, polyclonal rabbit antibody with the EnVision detection system (DAKO) was used. The slides were further developed using 3-amino-9-ethylcarbazole, counterstained with hemalaun and mount-

ed with glycerin gelatin. Stained sections were observed with an Axioplan 2 microscope equipped with an Axiocam digital camera. Staining intensity for  $\beta$ -catenin was scored as no staining (value 0), weak and fragmentary staining of cell membranes (value 1), moderate membrane staining of less than 50% of the tumor cells (value 2), moderate membrane staining of more than 50% of tumor cells (value 3) and strong membrane staining of more than 75% of tumor cells (value 4). The cyclin D1 staining was quantified as a percentage of

**Fig. 1.** Fzd5 protein expression in matched tumor/normal kidney samples. T = Tumor sample; N = matched normal sample. Sample identification numbers are given by the pathologist.



cyclin D1-immunoreactive nuclei in tumor cells in three fields (area: 18,641  $\mu\text{m}^2$ ) of each tumor sample. The total number of tumor nuclei ranged from 51 to 164. The correlation between Fzd5 and  $\beta$ -catenin staining, and between Fzd5 and cyclin D1 staining was evaluated by the Mann-Whitney U test.

## Results

### *Fzd mRNA Expression in Matched Human Tumor/Normal Tissue Samples*

Fzd expression in tumor tissue was compared with Fzd expression in matched normal tissue samples and normalized to the expression of the housekeeping gene mitochondrial ATP5b (table 2). A 3-fold increase was considered significant.

In the kidney tumor samples, in which 10 of 11 samples were clear cell carcinomas, Fzd5 was upregulated in 8 of the 11 samples. A similar observation was made for Fzd8 and Fzd2, which were upregulated in 7 and 5 renal tumor samples, respectively. None of the other Fzds showed consistent upregulation.

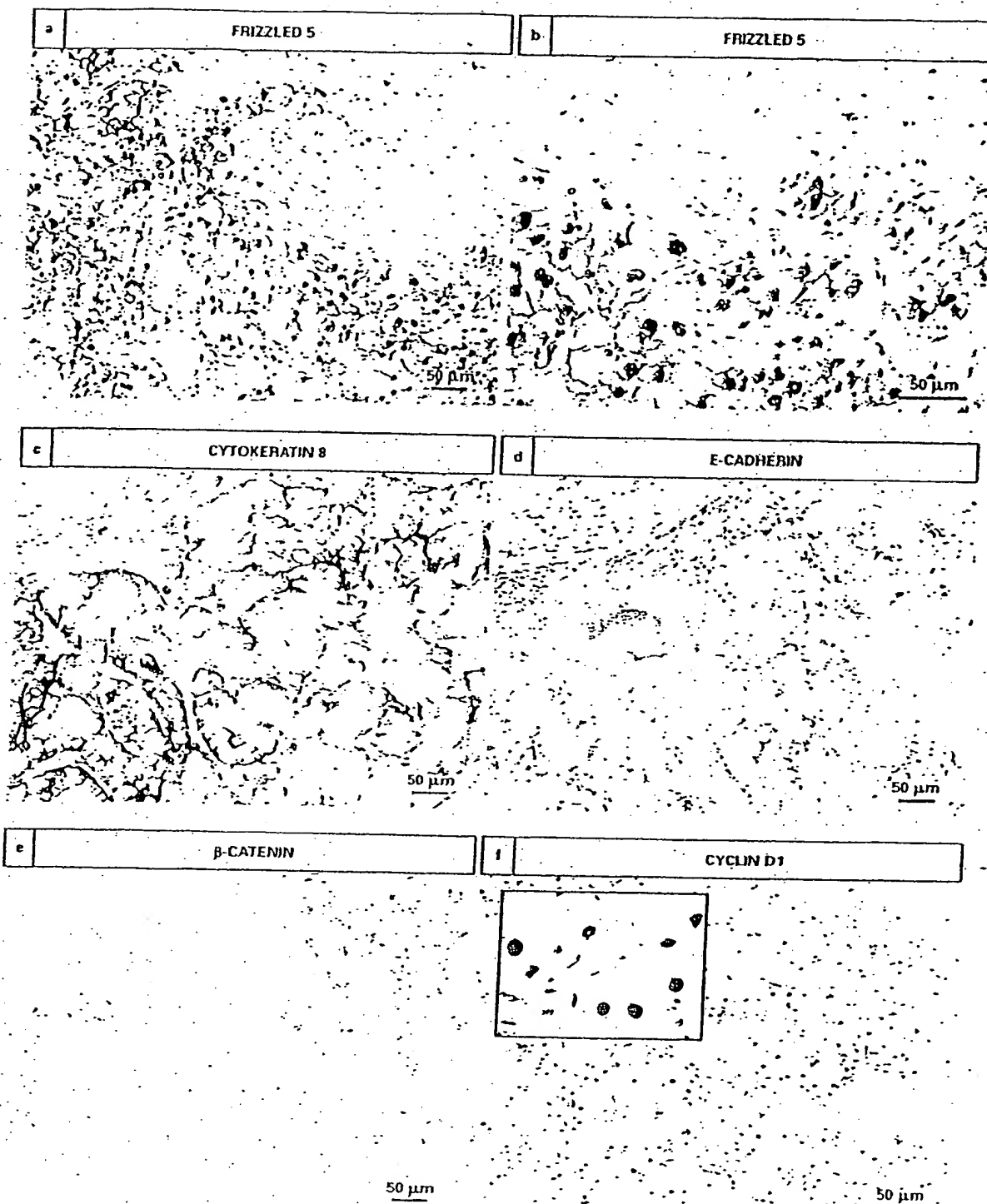
Both Fzd2 and Fzd3 were upregulated in 3 of 5 colon adenocarcinoma samples. No other Fzd expression was significantly different compared to the normal colon tissue sample. Fzd3 showed an increased expression in all 3 ovarian carcinoma samples. Fzd expression was not altered in any of the lung tumor samples. The Fzd expres-

sion level was observed to be relatively low in these lung tissues compared to the other tissues investigated.

### *Western Blot and Immunohistochemistry Analysis on Renal Carcinomas*

Western blotting was used to evaluate Fzd5 protein expression in the renal tissue samples used for mRNA expression analysis. Membrane fractions of the renal carcinoma and corresponding normal tissue samples were prepared. As previously shown (table 2), Fzd5 mRNA upregulation was detected in 8 of the 11 matched tumor/normal samples. Increased expression of Fzd5 protein was seen in membrane fractions from 9 of 11 samples (fig. 1). In most cases, concomitant increases in Fzd5 mRNA and protein levels were observed.

Hematoxylin-eosin staining of the cryosectioned tumors confirmed the presence of clear cell carcinoma. Fzd5 immunostaining in clear cell carcinoma (fig. 2a, b) was observed to be localized to cell membranes and to nuclei. Cytokeratin 8 (fig. 2c) and E-cadherin (fig. 2d) were also detected. E-cadherin labeling of cell membranes in clear cell carcinoma was less intense and patchy compared to epithelial cells of normal renal tissue.  $\beta$ -Catenin staining was confined to the cell membrane.  $\beta$ -Catenin levels in the clear cell carcinoma membranes were highly variable. Nuclear  $\beta$ -catenin staining was not observed in any of the samples. Epithelial cells in normal renal tissue showed intense membrane staining and some cytoplasmic



staining. In addition, weak  $\beta$ -catenin staining of endothelial cells was observed. A high number of cyclin D1-immunoreactive nuclei was observed in clear cell carcinoma (fig. 2f).

On the CL1 human kidney cancer tissue array, 30% ( $n = 9$ ) of the clear cell carcinoma tumor samples ( $n = 30$ ) showed Fzd5 immunoreactivity (fig. 3a). Membrane-associated  $\beta$ -catenin staining was observed in 33% of the Fzd5-positive tumor samples and 57% of Fzd5-negative clear renal cell carcinoma samples (table 3; fig. 3c, d). Again, nuclear  $\beta$ -catenin staining was never observed. Statistical analysis did not reveal a difference in the expression of  $\beta$ -catenin between Fzd5-positive and Fzd5-negative tumor samples (fig. 4a).

Nuclear cyclin D1 was observed in 89% of the Fzd5-positive clear cell carcinoma samples (table 3; fig. 3e). Only 38% of the Fzd5-negative clear cell carcinoma samples contained nuclear cyclin D1. Statistical analysis showed a significantly higher cyclin D1 expression in Fzd5-positive compared to Fzd5-negative tumor samples (fig. 4b).

#### *c-myc, Cyclin D1 and Peroxisome Proliferator-Activated Receptor $\delta$ Expression in Renal Carcinomas*

Wnt/ $\beta$ -catenin pathway activation in the kidney tissue samples was investigated looking at the expression of a number of target genes, which have previously been shown to be upregulated when the pathway is active. Gene expression of *c-myc*, cyclin D1 and peroxisome proliferator-activated receptor  $\delta$  (PPAR $\delta$ ) was analyzed. Increased expression of both *c-myc* and cyclin D1 genes have been implicated in cell proliferation, and carcinogenesis, and they represent two of the more important and closely studied target genes of the Wnt signaling pathway.

Expression of PPAR $\delta$  was investigated because it represents a direct target of the  $\beta$ -catenin pathway with T cell factor binding sites in its promoter. Expression of *c-myc* was found to be upregulated in 7 of 11, whilst cyclin D1 was upregulated in 10 of 11 kidney tumor samples (table 4). PPAR $\delta$  was upregulated in 9 cases. All three selected target genes showed a marked upregulation in the majority of renal tumors, which suggested that the Wnt/ $\beta$ -catenin pathway was activated in these samples.

#### **Discussion**

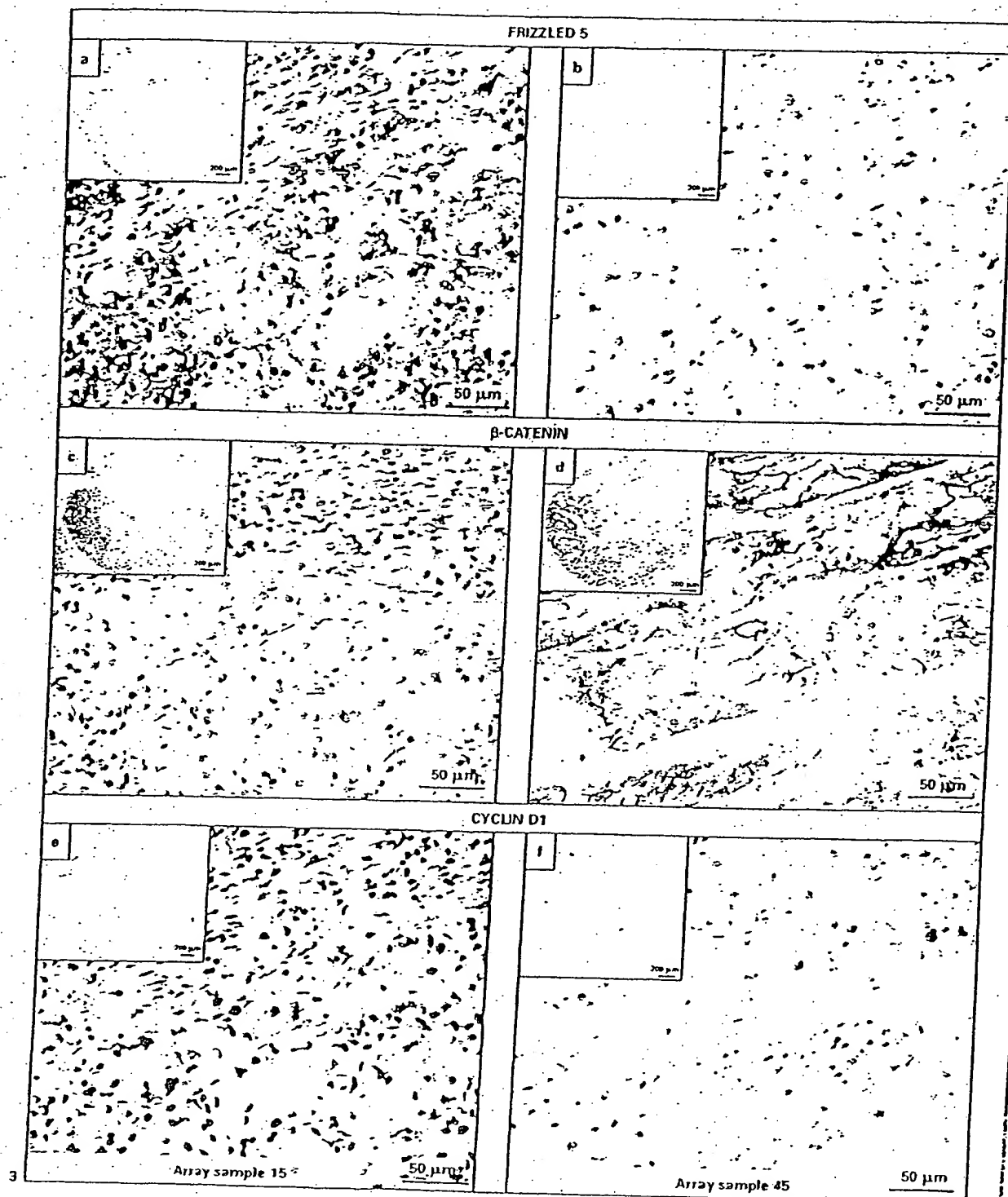
Fzd family member overexpression has been postulated to play key roles in different tumor types such as esophageal carcinoma [11], gastric cancer [12] and head and neck squamous cell carcinoma [13]. The current study evaluated the potential implication of Fzds as tumor-associated antigens in different tumor types. We screened a number of matched normal/tumor tissue samples for the expression of a variety of Fzds using real-time quantitative PCR.

Results obtained revealed that both Fzd5 and Fzd8 mRNA were overexpressed in the majority of renal carcinoma samples when compared to the matched normal kidney samples. Fzd2 and Fzd3 were upregulated in 3 of 5 colon adenocarcinoma samples. Fzd3 was also upregulated in the ovarian tumor tissue samples compared to the matched normal tissue samples. None of the other Fzds evaluated showed a specific differential expression pattern in any of the samples studied. Fzd5 and Fzd8 show 69.1% similarity and belong to the same subgroup of Fzds [14]. The significantly higher expression of Fzd5 and Fzd8 in the renal tumor samples, as compared to the normal renal samples, suggests a higher probability that this subgroup may be implicated in the progression of renal cancer. Therefore, we decided to further examine the possible role of Fzd5 in renal carcinoma.

We observed, using Western blotting, that protein levels were mostly consistent with mRNA levels in the tumor samples. In order to be able to determine the Fzd5 expression in a broader range of kidney tissues, we utilized a tissue array. Fzd5 membrane staining was detected in 9 of 30 (30%) clear cell carcinomas, and importantly, membrane staining was not detected in the matched 9 normal kidney tissue samples.

Since the Wnt signaling pathway appears to play an important role in embryonic development, in particular embryonic kidney induction [15, 16], activation of this pathway in the adult kidney due to mutation or overex-

**Fig. 2.** Distribution of Fzd5 (a, b), cytokeratin 8 (c), E-cadherin (d) and  $\beta$ -catenin (e) immunoreactivity in paraffin sections from a renal tumor processed by an alcohol fixative. From the same tumor, a formalin-fixed block was used for cyclin D1 immunostaining (f). Fzd5 immunostaining shows distinct immunoreactivity in cell membranes and in nuclei of clear cell renal carcinoma. Clear cells are immunoreactive for cytokeratin 8.  $\beta$ -Catenin and E-cadherin staining of membranes is rather weak, and not uniform, in clear cell renal carcinoma. Nuclear  $\beta$ -catenin immunoreactivity was not observed. In clear cell renal carcinoma, many nuclei showed cyclin D1 immunoreactivity. The inset in f shows a detailed view of the cyclin D1 labeling of nuclei in clear cell renal carcinoma.





**Table 3.** Correlation between Fzd5 and  $\beta$ -catenin or cyclin D1 expression

	Fzd5, %	
	+	-
$\beta$ -Catenin +	33	57
$\beta$ -Catenin -	67	43
Cyclin D1 +	89	38
Cyclin D1 -	11	62

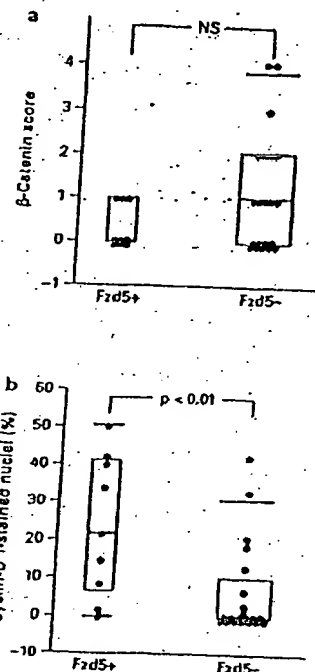
pression of one of the components of the pathway could be a determining factor in the development of renal cancers. Therefore, several studies have looked into the possible function the Wnt/ $\beta$ -catenin pathway plays in renal carcinogenesis. APC gene mutations have been demonstrated not to be involved in renal carcinoma [17, 18]. In addition,  $\beta$ -catenin mutations are rare events in renal carcinoma [19, 20]. Nevertheless, cytoplasmic accumulation of  $\beta$ -catenin has been reported in a number of renal cell carcinomas [19], and thus the Wnt signaling pathway.

**Table 4.** Wnt/ $\beta$ -catenin target gene mRNA expression in tumor samples

Sample <sup>a</sup>	Tissue	Tumor type	x-fold expression increase <sup>b</sup>		
			c-myc	cyclin D1	PPAR $\delta$
133702	kidney	adenocarcinoma	0.54	4.52	0.52
137770	kidney	renal cell carcinoma	13.9	28.91	5.53
138844	kidney	renal cell carcinoma	2.39	31.49	7.48
137146	kidney	renal cell carcinoma	7.62	15.38	3.15
137564	kidney	renal cell carcinoma	33.82	19.65	8.65
133408	kidney	renal cell carcinoma	7.8	8.92	4.86
139188	kidney	renal cell carcinoma	2.22	9.92	11.67
135699	kidney	renal cell carcinoma	12.18	22.73	5.53
139064	kidney	renal cell carcinoma	22.11	5.04	6.33
134585	kidney	renal cell carcinoma	1.79	1.14	1.67
140279	kidney	renal cell carcinoma	61.68	54.95	14.62

<sup>a</sup> Sample identification numbers were given by the pathologist.

<sup>b</sup> Results are expressed as x-fold increase of the gene in the tumor tissue sample compared to its matched normal tissue sample after normalizing both samples on the basis of their ATP5b content. A cutoff of 3-fold was used to define differential expression. Significant (>3-fold) increases in the expression level of the Fzd receptors are in italics.



**Fig. 3.** Fzd5,  $\beta$ -catenin and cyclin D1 immunostaining of the CL1 renal carcinoma tissue arrays. The left column of images represents serial sections from tumor sample 15. Insets show an overview of each tumor section on the serial tissue arrays: The Fzd5-immunoreactive clear cell renal carcinoma (a) of this tumor sample does not express  $\beta$ -catenin (c). Immunostaining for cyclin D1 (e) detects distinct labeling of nuclei in clear cell renal carcinoma. The right column of images is taken from serial sections of tumor sample 45. Clear cell renal carcinoma from this tumor sample does not express Fzd5 (b) and cyclin D1 (f) but does show distinct membrane  $\beta$ -catenin staining (d).

**Fig. 4.** Box plot charts (thick black line = median) illustrating the relationship between Fzd5 immunostaining and  $\beta$ -catenin (a) and cyclin D1 (b) expression in clear cell renal carcinoma. No significant correlation was observed between the  $\beta$ -catenin scores of Fzd5-positive and -negative clear cell renal carcinoma. Nuclear cyclin D1 staining in clear cell renal carcinoma showed a significant difference between Fzd5-positive and Fzd5-negative tumor samples.

might act as an inducer of tumorigenesis in the kidney. This view is supported by the observation that aberrant activity of the Wnt signaling pathway has been reported in renal-cancer-derived cell lines. Zang et al. [21] observed a higher expression level of Wnt5a and Fzd5 mRNA in the renal cancer cell line GRC-1 than in the normal renal cell line HK-2. Expression of  $\beta$ -catenin was also higher in GRC-1 than in HK-2.

To determine the status of the canonical Wnt signaling pathway in our renal carcinoma samples, we have quantitated the mRNA levels of three important target genes of T cell factor/lymphoid enhancer factor activation by  $\beta$ -catenin. The mRNA levels of these three target genes (*c-myc*, cyclin D1 and PPAR  $\delta$ ) correlated largely with the expression of Fzd5 in these samples, suggesting that the canonical pathway is activated. On the kidney tissue array, cyclin D1 protein expression showed a highly significant correlation with the Fzd5 expression in the tumor samples (table 3). Cyclin D1 protein is frequently overexpressed in various tumors, but in only a proportion of the cases is it due to amplification of the cyclin D1 gene [22]. Therefore, other mechanisms such as upregulation of gene transcription may play a substantial role in the overexpression of cyclin D1 [23–26]. Our data, showing increased cyclin D1 expression in renal carcinoma samples, are consistent with the results of Stassar et al. [27]. They studied genes that are associated with human renal carcinoma by suppression subtractive hybridization and reported 14 differentially expressed genes, including cyclin D1. Although we would have expected an increased nuclear  $\beta$ -catenin staining, nuclear accumulation of  $\beta$ -catenin was not observed in any of the tumors or on the tissue array.

This result is consistent with the data presented for renal cell carcinomas by Kim et al. [19]. They did not detect nuclear  $\beta$ -catenin staining in the 52 renal cell carcinomas examined. The lack of nuclear  $\beta$ -catenin staining has also been reported by others in tumors that might have arisen from Wnt/ $\beta$ -catenin pathway activation [28–31].

While expression of both Wnt5a and Fzd5 does induce duplication of the *Xenopus* head, exogenous expression of Fzd5 in a *Xenopus* model does not induce duplication of the head [32]. Fzd5 does not activate the  $\beta$ -catenin signaling pathway on its own, as the presence of its endogenous ligand is also required. Our results suggest that Fzd5 may have a role in renal cell carcinogenesis due to its frequent overexpression observed in these tumor samples, and we hypothesize that if Fzd5 is overexpressed, it has a rather limited effect on  $\beta$ -catenin signaling. However, in the presence of its endogenous still unknown ligand, it activates the canonical Wnt signaling pathway. The elucidation of this ligand and its binding characteristics is still under investigation. Ultimately, knowledge of the specific expression patterns of both Wnt and Fzd members could lead to directed tumor targeting or could be used as a tumor marker.

#### Acknowledgment

We are grateful to Prof. E. Van Marck (University of Antwerp) for providing us with the tissue samples.

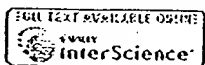
#### References

- Pandur P, Maurus D, Kuhl M: Increasingly complex: New players enter the Wnt signaling network. *Bioessays* 2002;24:881–884.
- Dale TC: Signal transduction by the Wnt family of ligands. *Biochem J* 1998;329:209–223.
- Oving JM, Clevers HC: Molecular causes of colon cancer. *Eur J Clin Invest* 2002;32:448–457.
- Pinson KI, Brennan J, Monkley S, Avery BJ, Skarnes WC: An LDL-receptor-related protein mediates Wnt signalling in mice. *Nature* 2000;407:535–538.
- Tamai K, Semenov M, Kato Y, Spokony R, Liu CM, Katsuyama Y, Hess F, Saint-Jeannet JP, He X: LDL-receptor-related proteins in Wnt signal transduction. *Nature* 2000;407:530–535.
- Miller JR: The Wnts. *Genome Biol* 2002; vol 3.
- Sato S, Daigo Y, Furukawa Y, Kato T, Miwa N, Nishiwaki T, Kawasoe T, Ishiguro H, Fujita M, Tokino T, Sasaki Y, Imaoka S, Murata M, Shimano T, Yamaoka Y, Nakamura Y: AXIN1 mutations in hepatocellular carcinomas, and growth suppression in cancer cells by virus-mediated transfer of AXIN1. *Nat Genet* 2000;24:245–250.
- Vider BZ, Zimmer A, Chastre E, Prevot S, Gerspach C, Estlin D, Wolloch Y, Tronick SR, Gazit A, Yaniv A: Evidence for the involvement of the Wnt 2 gene in human colorectal cancer. *Oncogene* 1996;12:153–158.
- Iozzo RV, Eichstetter I, Danielson KG: Aberrant expression of the growth factor Wnt-5A in human malignancy. *Cancer Res* 1995;55:3495–3499.
- Gerard CJ, Andrejka LM, Macina RA: Mitochondrial ATP synthase 6 as an endogenous control in the quantitative RT-PCR analysis of clinical cancer samples. *Mol Diagn* 2000;5:39–46.
- Tanaka S, Akiyoshi T, Mori M, Wands JR, Sugimachi K: A novel frizzled gene identified in human esophageal carcinoma mediates APC/beta-catenin signals. *Proc Natl Acad Sci USA* 1998;95:10164–10169.
- To KF, Chan MW, Leung WK, Yu J, Tong JH, Lee TL, Chan FK, Sung JJ: Alterations of frizzled (FZ3) and secreted frizzled related protein (sFRP) expression in gastric cancer. *Life Sci* 2001;70:483–489.
- Rhee CS, Sen M, Lu D, Wu C, Leoni L, Rubin J, Corr M, Carson DA: Wnt and frizzled receptors as potential targets for immunotherapy in head and neck squamous cell carcinomas. *Oncogene* 2002;21:6598–6605.

- 14 Saitoh T, Hirai M, Katoh M: Molecular cloning and characterization of human frizzled-8 gene on chromosome 10p11.2. *Int J Oncol* 2001;18:991-996.
- 15 Dressler GR: Tubulogenesis in the developing mammalian kidney. *Trends Cell Biol* 2003;12:390-395.
- 16 Vainio SJ, Itaranta PV, Perasaari JP, Uusitalo MS: Wnts as kidney tubule inducing factors. *Int J Dev Biol* 1999;43:419-423.
- 17 Bohm M, Wieland I, Stinhofer C, Otto T, Rubben H: Detection of loss of heterozygosity in the APC tumor suppressor gene in nonpapillary renal cell carcinoma by microdissection and polymerase chain reaction. *Urol Res* 1997;25:161-165.
- 18 Suzuki H, Ueda T, Komiya A, Okano T, Isaka S, Shimazaki J, Ito H: Mutational state of von Hippel-Lindau and adenomatous polyposis coli genes in renal tumors. *Oncology* 1997;54:252-257.
- 19 Kim YS, Kang YK, Kim JB, Han SA, Kim KJ, Paik SR: Beta-catenin expression and mutational analysis in renal cell carcinomas. *Pathol Int* 2000;50:725-730.
- 20 Ueda M, Gemmill RM, West J, Winn R, Sugita M, Tanaka N, Ueki M, Drabkin HA: Mutations of the  $\beta$ - and  $\gamma$ -catenin genes are uncommon in human lung, breast, kidney, cervical and ovarian carcinomas. *Br J Cancer* 2001;85:64-68.
- 21 Zang T, Zhuang L, Zhang Z, Xin D, Guo Y: Aberrant activity of WNT/frizzled signaling pathway in renal cancer cell lines. *Chin Sci Bull* 2000;45:1703-1707.
- 22 Ozturk M: Genetic aspects of hepatocellular carcinogenesis. *Semin Liver Dis* 1999;19:235-242.
- 23 Aktas H, Cai H, Cooper GM: Ras links growth factor signaling to the cell cycle machinery via regulation of cyclin D1 and the Cdk inhibitor p27KIP1. *Mol Cell Biol* 1997;17:3850-3857.
- 24 Weber JD, Raben DM, Phillips PJ, Baldassare JJ: Sustained activation of extracellular-signal-regulated kinase 1 (ERK1) is required for the continued expression of cyclin D1 in G1 phase. *Biochem J* 1997;326:61-68.
- 25 Lavoie JN, L'Allemain G, Brunet A, Muller R, Pouyssegur J: Cyclin D1 expression is regulated positively by the p42/p44MAPK and negatively by the p38/HOGMAPK pathway. *J Biol Chem* 1996;271:20608-20616.
- 26 Treinies I, Patterson HF, Hooper S, Wilson R, Marshall CJ: Activated MEK stimulates expression of AP-1 components independently of phosphatidylinositol 3-kinase (PI3-kinase) but requires a PI3-kinase signal to stimulate DNA synthesis. *Mol Cell Biol* 1999;19:321-329.
- 27 Stassar MJ, Devitt G, Brosius M, Rinnab L, Prang J, Schradin T, Simon J, Petersen S, Kopp-Schneider A, Zoller M: Identification of human renal cell carcinoma associated genes by suppression subtractive hybridization. *Br J Cancer* 2001;85:1372-1382.
- 28 Wong SC, Lo SF, Lee KC, Yam JW, Chan JK, Wendy Hsiao WL: Expression of frizzled-related protein and Wnt-signalling molecules in invasive human breast tumours. *J Pathol* 2002;196:145-153.
- 29 Ueda T, Ikeguchi M, Hirooka Y, Kaibara N, Terada T: Beta-catenin and cyclin D1 expression in human hepatocellular carcinoma. *Oncol Rep* 2002;9:1197-1203.
- 30 Tanaka Y, Kato K, Notohara K, Nakatani Y, Miyake T, Ijiri R, Nishimata S, Ishida Y, Kigasawa H, Ohama Y, Tsukayama C, Kobayashi Y, Horie H: Significance of aberrant (cytoplasmic/nuclear) expression of beta-catenin in pancreaticoblastoma. *J Pathol* 2003;199:185-190.
- 31 Qiao Q, Ramadani M, Gansauge S, Gansauge F, Leder G, Beger HG: Reduced membranous and ectopic cytoplasmic expression of beta-catenin correlate with cyclin D1 overexpression and poor prognosis in pancreatic cancer. *Int J Cancer* 2001;95:194-197.
- 32 He X, Saint-Jeannet JP, Wang Y, Nathans J, Dawid I, Varmus H: A member of the frizzled protein family mediating axis induction by Wnt-5A. *Science* 1997;275:1652-1654.

Erratum in:

- Int J Cancer 2002 Feb 20;97(6):878.



**Immunohistochemical analysis of NY-ESO-1 antigen expression in normal and malignant human tissues.**

Jungbluth AA, Chen YT, Stockert E, Busam KJ, Kolb D, Iversen K, Coplan K, Williamson B, Altorki N, Old LJ.

Ludwig Institute for Cancer Research, Memorial Sloan-Kettering Cancer Center, New York, NY, USA. [jungblua@mskcc.org](mailto:jungblua@mskcc.org)

NY-ESO-1, a member of the CT (cancer/testis) family of antigens, is expressed in normal testis and in a range of human tumor types. Knowledge of NY-ESO-1 expression has depended on RT-PCR detection of mRNA and there is a need for detecting NY-ESO-1 at the protein level. In the present study, a method for the immunochemical detection of NY-ESO-1 in paraffin-embedded tissues has been developed and used to define the expression pattern of NY-ESO-1 in normal tissues and in a panel of human tumors. No normal tissue other than testis showed NY-ESO-1 reactivity, and expression in testis was restricted to germ cells particularly spermatogonia. In human tumors, the frequency of NY-ESO-1 antigen expression corresponds with past analysis of NY-ESO-1 mRNA expression e.g., 20-30% of lung cancers, bladder cancers and melanoma, and no expression in colon and renal cancer. Co-typing of NY-ESO-1 antigen and mRNA expression in a large panel of lung cancers showed a good correlation. There is great variability in NY-ESO-1 expression in individual tumors, ranging from an infrequent homogeneous pattern of staining to highly heterogeneous antigen expression. Copyright 2001 Wiley-Liss, Inc.

PMID: 11351307 [PubMed - indexed for MEDLINE]

Full text article at  
[www.biolreprod.org](http://www.biolreprod.org)

**Multidrug resistance phosphoglycoprotein (ABCB1) in the mouse placenta: fetal protection.**

Kalabis GM, Kostaki A, Andrews MH, Petropoulos S, Gibb W, Matthews SG.

Department of Physiology, University of Toronto, Ontario, Canada.

The multidrug resistance phosphoglycoprotein ATP-binding cassette subfamily B (ABCB1) actively extrudes a range of structurally and functionally diverse xenobiotics as well as glucocorticoids. ABCB1 is present in many cancer cell types as well as in normal tissues. Although it has been localized within the mouse placenta, virtually nothing is known about its regulation. In the mouse, two genes, *Abcb1a* and *Abcb1b*, encode ABCB1. We hypothesized that there are changes in placental *Abcb1a* and *Abcb1b* gene expression and ABCB1 protein levels during pregnancy. Using *in situ* hybridization, we demonstrated that *Abcb1b* mRNA is the predominant placental isoform and that there are profound gestational changes in the expression of both *Abcb1a* and *Abcb1b* mRNA. Placentas from pregnant mice were analyzed between Embryonic Days (E) 9.5 and 19 (term approximately 19.5d). *Abcb1b* mRNA was detected in invading trophoblast cells by E9.5, peaked within the placental labyrinth at E12.5, and then progressively decreased toward term ( $P < 0.0001$ ). *Abcb1a* mRNA, although lower than that of *Abcb1b* at midgestation, paralleled changes in *Abcb1b* mRNA. Changes in *Abcb1* mRNA were reflected by a significant decrease in ABCB1 protein ( $P < 0.05$ ). A strong correlation existed between placental *Abcb1b* mRNA and maternal progesterone concentrations, indicating a potential role of progesterone in regulation of placental *Abcb1b* mRNA. In conclusion, there are dramatic decreases in *Abcb1a* and *Abcb1b* mRNA and in ABCB1 at the maternal-fetal interface over the second half of gestation, suggesting that the fetus may become increasingly susceptible to the influences of xenobiotics and natural steroids in the maternal circulation.

PMID: 15917342 [PubMed - in process]

**Expression of human telomerase reverse transcriptase gene and protein, and of estrogen and progesterone receptors, in breast tumors: Preliminary data from neo-adjuvant chemotherapy.**

Kamori M, Izumiyama N, Hashimoto M, Nakamura K, Okano T, Kurabayashi R, Naoki H, Honma N, Ogawa T, Kaminishi M, Takubo K.

Division of Breast and Endocrine Surgery, Department of Surgery, Graduate School of Medicine, The University of Tokyo, Bunkyo-ku, Tokyo 113-8655, Japan. kamori-dis@umin.ac.jp.

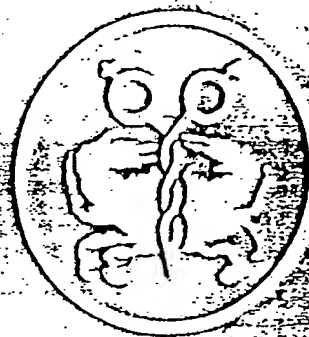
Human telomerase reverse transcriptase (hTERT), the catalytic subunit of telomerase, is very closely associated with telomerase activity. Telomerase has been implicated in cellular immortalization and carcinogenesis. In situ detection of hTERT will aid in determining the localization of telomerase-positive cells. The aim of this study was to detect expression of hTERT mRNA, hTERT protein, estrogen receptor (ER) and progesterone receptor (PR) in paraffin-embedded breast tissue samples and to investigate the relationship between hTERT expression and various clinicopathological parameters in breast tumorigenesis. We used in situ hybridization (ISH) to examine hTERT gene expression, and immunohistochemistry (IHC) to examine expression of hTERT protein, ER and PR, in breast tissues including 64 adenocarcinomas, 2 phyllode tumors and their adjacent normal breast tissues. hTERT gene expression was detected by ISH in 56 (88%) carcinomas, but in neither of the 2 phyllode tumors. hTERT protein expression was detected by IHC in 52 (81%) carcinomas, but in neither of the 2 phyllode tumors. Moreover, ER and PR were expressed in 42 (66%) and 42 (66%) carcinomas, respectively, and in neither of the 2 phyllode tumors. In 4 cases of breast carcinoma that strongly expressed hTERT gene and protein before treatment, neo-adjuvant chemotherapy led to disappearance of gene and protein expression in all cases. There was a strong correlation between detection of hTERT gene expression by ISH and of hTERT protein by ICH in tissue specimens from breast tumors. These results suggest that detection of hTERT protein by ICH can be used to distinguish breast cancers as a potential diagnostic and therapeutic marker.

PMID: 16211220 [PubMed - in process]

181  
v. 27  
no. 5  
2005 Nov

Received on: 05-10-20  
International journal of  
oncology.

# Journal of Oncology



ISSN 1070-6139

An international journal devoted to Oncology Research and Cancer Treatment

VOLUME 27, NUMBER 5, NOVEMBER 2005



# Expression of human telomerase reverse transcriptase gene and protein, and of estrogen and progesterone receptors, in breast tumors: Preliminary data from neo-adjuvant chemotherapy

MAKOTO KAMMORI<sup>1,2</sup>, NAOTAKA IZUMIYAMA<sup>2</sup>, MASANORI HASHIMOTO<sup>1</sup>, KEN-ICHI NAKAMURA<sup>2</sup>,  
TADAO OKANO<sup>3</sup>, RIE KURABAYASHI<sup>1</sup>, HIKI NAOKI<sup>1</sup>, NAOKO HONMA<sup>2</sup>, TOSHIHISA OGAWA<sup>1</sup>,  
MICHIO KAMINISHI<sup>1</sup> and KAIYO TAKUBO<sup>2</sup>

<sup>1</sup>Division of Breast and Endocrine Surgery, Department of Surgery, Graduate School of Medicine,  
The University of Tokyo; <sup>2</sup>Human Tissue Research Group, Tokyo Metropolitan Institute of Gerontology;

<sup>3</sup>Tokyo Metropolitan Tama Cancer Detection Center, Tokyo, Japan

Received January 25, 2005; Accepted March 24, 2005

**Abstract.** Human telomerase reverse transcriptase (hTERT), the catalytic subunit of telomerase, is very closely associated with telomerase activity. Telomerase has been implicated in cellular immortalization and carcinogenesis. *In situ* detection of hTERT will aid in determining the localization of telomerase-positive cells. The aim of this study was to detect expression of hTERT mRNA, hTERT protein, estrogen receptor (ER) and progesterone receptor (PR) in paraffin-embedded breast tissue samples and to investigate the relationship between hTERT expression and various clinicopathological parameters in breast tumorigenesis. We used *in situ* hybridization (ISH) to examine hTERT gene expression, and immunohistochemistry (IHC) to examine expression of hTERT protein, ER and PR, in breast tissues including 64 adenocarcinomas, 2 phyllode tumors and their adjacent normal breast tissues. hTERT gene expression was detected by ISH in 56 (88%) carcinomas, but in neither of the 2 phyllode tumors. hTERT protein expression was detected by IHC in 52 (81%) carcinomas, but in neither of the 2 phyllode tumors. Moreover, ER and PR were expressed in 42 (66%) and 42 (66%) carcinomas, respectively, and in neither of the 2 phyllode tumors. In 4 cases of breast carcinoma that strongly expressed hTERT gene and protein before treatment, neo-adjuvant chemotherapy led to disappearance of gene and protein expression in all cases. There was a strong correlation between detection of hTERT gene expression by ISH and of hTERT protein by ICH in tissue specimens from breast tumors. These results suggest that detection of hTERT

protein by ICH can be used to distinguish breast cancers as a potential diagnostic and therapeutic marker.

## Introduction

Breast cancer is the most frequent malignancy in women, affecting up to one in every eight females worldwide. The most important clinicopathological prognostic parameter so far identified is the absence or presence of lymph node metastasis, but the identification of further parameters would facilitate an individually based risk-directed therapy (1). A promising emerging molecular marker is telomerase, a ribonucleoprotein enzyme complex, which when activated or upregulated allows tumor cells to escape from cellular senescence and to proliferate indefinitely (2). The human telomere is a simple repeat sequence of six bases (TTAGGG) that is located at the ends of each chromosome (3). Telomeres are believed to protect against degeneration, reconstruction, fusion, and loss (4) and to promote the homologous pairing of chromosomes (5). The end-to-end chromosome fusions observed in some tumors may result from the loss of telomeres and may be partly responsible for the genetic instability associated with tumorigenesis. Telomerase catalyzes the synthesis of telomere DNA and facilitates cell immortalization through the stabilization of chromosomal structure (6-8). Although the expression of the human RNA component of telomerase (hTERC) is widespread, the restricted expression pattern of the mRNA of hTERT, the human telomerase catalytic subunit gene, is correlated with telomerase activity (8-13). As telomerase activity seems to be the key player in tumor cell immortality, it has importance as a target molecule for anti-cancer therapy. Telomerase activity has been shown to correlate with poor clinical outcome in neuroblastomas and other tumors (14). For breast cancer, however, telomerase activity is a controversial prognostic marker: some studies suggest that telomerase activity, clinicopathological parameters and disease outcome are linked, whereas others do not find this association (14-23).

**Correspondence to:** Dr Makoto Kammori, Division of Breast and Endocrine Surgery, Department of Surgery, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan  
E-mail: kammori-dis@umin.ac.jp

**Key words:** human telomerase reverse transcriptase, telomerase, estrogen receptor, breast cancer, chemotherapy



We have succeeded in very clearly and sensitively demonstrating hTERT mRNA in thyroid, colorectal, parathyroid and lung tissues by use of an oligonucleotide probe (13,24-26). Strong correlation has been observed between hTERT mRNA and/or protein expression and telomerase in a variety of malignant tumors (13,14,24,25,27,28). In the present study, we used ISH to examine expression of the hTERT gene, and IHC to examine expression of hTERT protein, ER and PR, in 64 carcinomas and 2 phyllodes tumors of breast to determine whether hTERT protein can be used to differentiate breast cancers. We also analyzed hTERT mRNA and protein expression with special reference to clinical features and histological findings to investigate the potential role of hTERT mRNA expression analysis in predicting the biological characteristics of breast cancers. Since hTERT expression in breast tumors has not previously been analysed by ISH or IHC, our investigation also examined various clinicopathological parameters, including age, histopathological type, tumor size, lymph node status, relapses, and the expression of ER and PR.

#### Materials and methods

**Tissue collection.** Sixty-six samples were obtained during 66 mastectomies: 64 breast carcinomas, 2 phyllodes tumors and 66 specimens of the adjacent normal breast gland. In 4 cases, samples were obtained during core needle biopsies (CNB) before neo-adjuvant therapy, and again during mastectomies after neo-adjuvant therapy; for these cases, all measurements and examinations were performed both before and after the neo-adjuvant therapy. The patients ranged in age from 32 to 90 years. The patients with carcinomas ranged in age from 37 to 90, mean 56, years and were all women. The women with phyllode tumors were aged 32 and 38 years, respectively. The surgical and CNB tissue samples were frozen rapidly with liquid nitrogen and stored at -80°C until fixation. Then, they were fixed in 10% buffer formalin solution and embedded in paraffin. Surgical and CNB samples were collected from the patients after obtaining their informed consent, and the study protocol was approved by the Medical Department of the University of Tokyo Ethics Committee. The pathologic diagnoses were made by the surgical pathological specialists at our institute on the basis of examination of hematoxylin-eosin stained slides. A pathological review was performed for all breast tumors according to the BRE score. pT and pN staging were assigned according to the 1997 WHO classification (7th edition).

MCF-7 human breast cancer cells, kindly provided by the Cell Resource Center for Biomedical Research, Institute of Development, Aging and Cancer, Tohoku University, were used as positive controls. The cells were incubated in RPMI-1640 medium with 25 mM HEPES buffer, L-glutamine, and 10% fetal bovine serum (Gibco, Grand Island, NY) on a chamber-attached slideglass (Lab-Tek® Chamber Slide™; Nalge Nunc International, Naperville, IL) in a humidified 5% CO<sub>2</sub> atmosphere at 37°C. The cells were then fixed with 10% buffered neutral formalin (Sigma Chemical Co., St. Louis, MO). The cultured MCF-7 cell line that was used as a positive control was tested for telomerase with a PCR-based standard TRAP assay (6,13). These cells were also used to prepare cell blocks. Briefly, the cells were fixed in 10% buffered neutral

formalin, resuspended in molten agarose and then embedded in paraffin. Sections from these cell blocks were used as positive procedural controls in ISH and IHC. The negative control in ISH was obtained by replacing the oligonucleotide probe with RNase. The negative control in IHC was obtained by replacing the primary antibody with Tris-buffered saline (TBS).

**Oligonucleotide probe for ISH.** The specificity of the oligonucleotide sequence was initially determined by a GenEMBL database search using the Genetics Computer Group Sequence Analysis Program (GCG, Madison, WI) based on the fastA algorithm (29); the sequence exhibited 100% homology with the hTERT gene sequence. A d(T)<sub>20</sub> oligonucleotide was used to verify the integrity and lack of degradation of the mRNA in each sample. All oligonucleotide probes were synthesized with a hapten-labeled nucleotide, such as digoxigenin-dUTP (Boehringer-Mannheim), at the 3' end via direct coupling by using standard phosphoramidite chemistry (Research Genetics, Huntsville, AL) (30). The probe used for detection of hTERT by ISH was generated from the original sequence for *Homo sapiens* telomerase reverse transcriptase (AF015950; 2766-2800: 5'-GCCTCGTCTTCTACAGGA AGTTCACCACTGTCTT-3' (13,24-26).

**ISH.** ISH was performed with the GenPoint nucleic acid hyper-detection system (Dako, Carpinteria, CA) (31). Formalin-fixed, paraffin-embedded tissue sections (5 µm thick) were deparaffinized in xylene and a graded alcohol series. Tissue sections and CNB samples were then pretreated with target retrieval solution (Dako, S1700) at 95°C and proteinase K (Dako, S3004) at room temperature. Next, the tissues and CNB samples were fixed in 0.3% hydrogen peroxide followed by a methyl alcohol series at room temperature. Digoxigenin-labeled anti-sense oligonucleotide in mRNA *in situ* hybridization solution (Dako, S3304) was placed over the tissues and CNB samples. After hybridization at 37°C overnight, the slides were washed in stringent wash solution (Dako, GenPoint System Kit) at 45°C. The tissues and CNB samples were exposed to avidin blocking solution (Dako, X0590) at room temperature followed by biotin blocking solution (Dako, X0590) at room temperature. The tissues and samples were then incubated at room temperature with a sheep monoclonal hapten-labeled anti-digoxigenin antibody (Dako, p5104), and the slides were then fixed with biotinyl tyramine (Dako, GenPoint System Kit) at room temperature. Finally, the slides were incubated with HRP-conjugated streptavidin (Dako, GenPoint System Kit) at room temperature. Since 3,3'-diaminobenzidine tetrahydrochloride (DAB) was used as the substrate, a positive reaction was visible as a brown color under a light microscope. The sections were weakly counter-stained with 0.1% hematoxylin.

**IHC.** IHC was performed by the avidin-biotin complex/horseradish peroxidase method. Tissue sections were stained for hTERT with a commercially available monoclonal antibody (NCL-L-hTERT; Novocastra, Newcastle upon Tyne, UK). Sections were dewaxed in xylene. Antigen retrieval was done by incubating sections immersed in 0.01 M citrate buffer at pH 6.0 in a microwave oven at 99°C. The sections were allowed to cool down at room temperature. The sections

MCF-7 Lysis Case 6 Case 14 Case 23 Case 44 Case 59

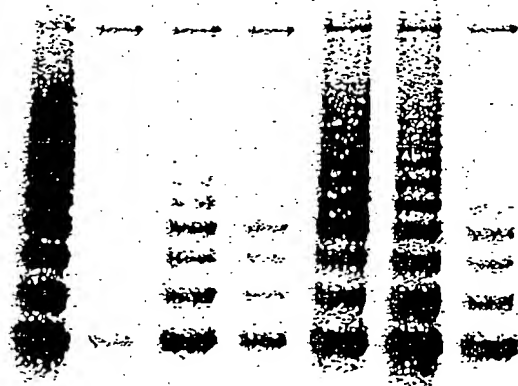


Figure 1. Representative results of the TRAP assay. If functional telomerase is present, the enzyme adds DNA to the substrate in 6-base-pair (bp) increments, resulting in a ladder-like distribution of products. The 6-bp ladder signals are apparent for MCF-7 and breast cancers (case nos. 6, 14, 23, 44 and 59) and are not apparent for lysis buffer as the negative control. An extract of MCF-7 was used as a positive control for the TRAP assay and as an Internal Telomerase Assay Standard (ITAS) positive control for PCR amplification, with lysis buffer as the negative control (Lysis).

were then immersed in 1% hydrogen peroxide ( $H_2O_2$ ) in methanol to block endogenous peroxidase activity. Following that, the sections were washed in TBS (pH 7.6) before being incubated in normal rabbit serum for 20 min to block non-specific binding. After draining off the excess serum, the sections were incubated with the primary antibody at room temperature. The sections were washed in TBS before being incubated with the secondary antibody (biotinylated rabbit anti-mouse, Dako). The sections were washed again with TBS and incubated with avidin-biotin complex/horseradish peroxidase. After washing the sections with TBS, peroxidase activity was visualized under light microscopy by applying DAB chromogen (Dako). The sections were counterstained with hematoxylin, dehydrated in increasing grades of alcohol and finally mounted in dibutyl phthalate (DPX) mountant.

Homogeneous staining or a speckled/dotted pattern in the nucleus was considered positive staining, and absence of distinct nuclear staining was taken as negative staining. Grading of the percentage of stained cells (hTERT labeling index) was performed by previously published criteria (1) as follows: Grade 1, negative staining; Grade 2, 1-10% positive staining nuclei; Grade 3, 11-50% positive nuclei; and Grade 4, >50% positive nuclei. Immunostained slides for ER and PR were scored as previously described (32,33). In brief, each entire slide was evaluated by light microscopy. First, a proportion score was assigned, which represents the estimated proportion of positive-staining tumor cells (0, none; 1, <1/100; 2, 1/100 to 1/10; 3, 1/10 to 1/3; 4, 1/3 to 2/3; and 5, >2/3). Next, an intensity score was assigned, which represents the average intensity of positive tumor cells (0, none; 1, weak; 2, intermediate; and 3, strong). The proportion and intensity scores were then added to obtain a total score, which ranged from 0 to 8. Slides were scored by pathologists who did not have knowledge of ligand-binding results or patient outcome.

Table I. Relationships between mRNA status (negative/positive) by ISH and standard clinical, pathological, and biological factors in the 66 tumors.

	Total population (%)	No. of patients (%)		P-value <sup>a</sup>
		hTERT negative	hTERT positive	
Total	66	10 (15.2)	56 (84.8)	
Age				NS
≤50	26	6 (23.1)	20 (76.9)	
>50	40	6 (15.0)	34 (85.0)	
Histopathological type				NS
Scirrhous	32	2 (6.4)	30 (93.6)	
Papillotubular	20	2 (10.0)	18 (90.0)	
Solid tubular	6	1 (16.7)	5 (83.3)	
Mucinous	2	1 (50.0)	1 (50.0)	
Non-invasive	4	2 (50.0)	2 (50.0)	
Phyllodes	2	2 (100)	0 (0)	
Tumor size (cm) <sup>b</sup>				NS
T1 (<2.0)	20	0 (0)	20 (100)	
T2 (2.0-5.0)	34	6 (17.6)	28 (82.4)	
T3 (>5.0)	10	2 (20.0)	8 (80.0)	
Lymph node status <sup>b</sup>				NS
pN0	34	6 (17.6)	28 (82.4)	
pN1	26	2 (7.7)	24 (92.3)	
pN2+pNM	4	0 (0)	4 (100)	
Relapse				NS
+	10	0 (0)	10 (100)	
-	56	10 (17.6)	46 (82.4)	
ER expression				NS
+ (≥2)	42	6 (14.3)	36 (85.7)	
- (<2)	24	4 (16.7)	20 (83.3)	
PR expression				NS
+ (≥2)	42	4 (9.5)	38 (90.5)	
- (<2)	24	6 (25.0)	18 (75.0)	

<sup>a</sup>χ<sup>2</sup> test. NS, not significant. <sup>b</sup>Information available for 64 patients.

**Statistical analysis.** Differences in p-values were analyzed with the χ<sup>2</sup> test for independence, and Fisher's test was used for correlations. In all comparisons, p<0.05 was considered significant.

## Results

Representative results of the TRAP assay are shown in Fig. 1. The cultured cells, which were tested for telomerase activity

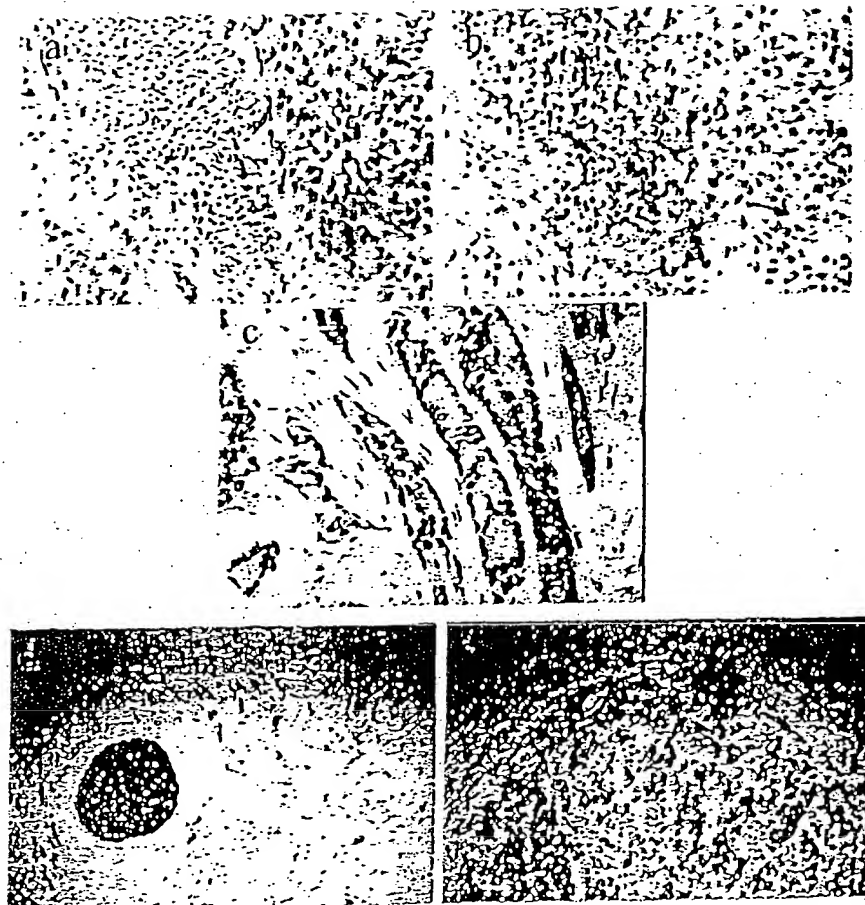


Figure 2. Correlation between histologic diagnosis, human telomerase reverse transcriptase (hTERT) mRNA by ISH, hTERT protein by IHC, estrogen receptor (ER) by IHC and progesterone receptor (PR) by IHC in breast cancers. (a, H&E); (b, hTERT mRNA); (c, hTERT protein); (d, ER) and (e, PR).

with the TRAP assay, gave positive results with all procedural controls (MCF-7 and 5 breast cancer samples) (Fig. 1).

ISH revealed that hTERT mRNA was strongly expressed in the nuclei and cytoplasm of almost all of the MCF-7 human cancer cells (data not shown). Expression of hTERT mRNA was detected in 56 (88%) of the 64 breast cancers and in none of the phyllodes tumors of the breast (Table I) with the anti-sense probe, whereas no expression was detected with the anti-sense probe treated with RNase (data not shown). The levels of expression were heterogeneous within the carcinomatous regions. Strong expression of hTERT mRNA was not confined to the carcinomatous regions but was also detected in infiltrating lymphocytes (Fig. 2a and b). Higher expression levels of both signals of hTERT mRNA were detected in some sections containing both carcinomas and lymphocytes, compared with the adjacent non-cancerous mammary gland, but no clear differences in signal intensity were observed between carcinomas and lymphocytes. The signals in both the normal and cancer tissues were mainly present in the lymphocytes, and the signal intensity was similar in both, although a precise quantitative comparison of the *in situ* signals was impossible.

IHC revealed that hTERT protein was strongly expressed in the nuclei, nuclear membrane and cytoplasm of almost all of the MCF-7 human cancer cells (data not shown). Expression of hTERT protein was detected in 52 (81%) of the 64 breast cancers and in none of the phyllode tumors of the breast (Table II). The levels of expression were heterogeneous within the carcinomatous regions. As shown in Fig. 2c, strong expression of hTERT protein was observed in nuclei, nuclear membrane and cytoplasm, similar to the pattern in MCF-7 human cancer cells. Normal mammary gland and stromal cells generally showed negative immunoreactivity against hTERT protein antibody.

A nuclear signal for the ER (Fig. 2d), as assessed by IHC, was observed in 36 (56%) of the 64 breast cancers and in none of the phyllode tumors of the breast, with positive scores ranging from 2 to 8 (Tables I and II). A nuclear signal for the PR (Fig. 2e), as assessed by IHC, was observed in 38 (59%) of the 64 breast cancers and in none of the phyllode tumors of the breast, with positive scores ranging from 2 to 8 (Tables I and II).

We used ISH and IHC to examine hTERT expression in 4 cases of breast cancer before and after neo-adjuvant

Table II. Relationships between mRNA status (negative/positive) by IHC and standard clinical, pathological, and biological factors in the 66 tumors.

	Total	No. of patients (%)		P-value <sup>a</sup>
		hTERT negative	hTERT positive	
Total	66	14 (21.2)	52 (78.8)	
Age				NS
≤50	26	10 (38.5)	16 (61.5)	
>50	40	4 (10.0)	36 (90.0)	
Histopathological type				NS
Scirrhous	32	2 (6.4)	30 (93.6)	
Papillotubular	20	4 (20.0)	16 (80.0)	
Solid tubular	6	4 (66.7)	2 (33.3)	
Mucinous	2	0 (0)	2 (100)	
Non-invasive	4	2 (50.0)	2 (50.0)	
Phyllodes	2	2 (100)	0 (0)	
Tumor size (cm) <sup>b</sup>				NS
T1 (<2.0)	20	4 (20.0)	16 (80.0)	
T2 (2.0-5.0)	34	6 (17.6)	28 (82.4)	
T3 (>5.0)	10	2 (20.0)	8 (80.0)	
Lymph node status <sup>b</sup>				NS
pN0	34	6 (17.6)	28 (82.4)	
pN1	26	6 (23.1)	20 (76.9)	
pN2+pNM	4	0 (0)	4 (100)	
Relapse				NS
+	10	1 (10.0)	9 (90.0)	
-	56	13 (23.2)	43 (79.6)	
ER expression				NS
+ (≥2)	42	6 (14.3)	36 (85.7)	
- (<2)	24	8 (33.3)	16 (66.7)	
PR expression				NS
+ (≥2)	42	6 (14.3)	36 (85.7)	
- (<2)	24	8 (33.3)	16 (66.7)	

<sup>a</sup> $\chi^2$  test. NS, not significant. <sup>b</sup>Information available for 64 patients.

chemotherapy. Before chemotherapy, all 4 of the breast carcinomas strongly expressed hTERT by both ISH and IHC. After chemotherapy, hTERT expression completely disappeared in all 4 cases (Table III). hTERT expression by lymphocytes was detectable by ISH and IHC both before and after chemotherapy in all 4 cases, and the level of expression did not appear to be altered by treatment.

No correlation was observed between hTERT mRNA expression and any of the clinicopathological parameters: age, histopathological type, tumor size, lymph node status,

Table III. Relationship of hTERT mRNA and protein expression before and after neoadjuvant chemotherapy.

Case	Age	Neoadjuvant	hTERT mRNA		hTERT protein	
			Before	After	Before	After
1	80	Anastrozole	+	-	+	-
2	78	Anastrozole	+	-	+	-
3	35	FEC <sup>a</sup>	+	-	+	-
4	37	AC <sup>b</sup>	+	-	+	-

<sup>a</sup>FEC, 5FU (500 mg/m<sup>2</sup>), Epirubicin (70 mg/m<sup>2</sup>), Cyclophosphamide (500 mg/m<sup>2</sup>). <sup>b</sup>AC, Doxorubicin (60 mg/m<sup>2</sup>), Cyclophosphamide (500 mg/m<sup>2</sup>). Before, before neoadjuvant chemotherapy. After, after neoadjuvant chemotherapy. +, positive; -, negative.

relapses, and the expression of ER and PR. Similarly, there was no correlation between hTERT protein expression and any of these clinicopathological parameters. There was a correlation between hTERT mRNA expression and hTERT protein expression in breast cancers ( $p < 0.005$ ).

#### Discussion

This study reports a comparison of hTERT mRNA expression by ISH and hTERT protein expression by IHC in tissue sections from breast tumors. hTERT mRNA was detected by ISH in 56 of the 64 breast cancers and in MCF-7 human breast cancer cells. Breast cancer cell nuclei stained strongly positive with the specific anti-sense probe but not with the anti-sense probe treated with RNase (data not shown). Tissue lymphocytes also stained positively with the anti-sense probe, but the stromal cells did not. Expression of hTERT protein was observed by IHC in 52 of the 64 breast cancers. hTERT mRNA and protein expressions were highly correlated in breast cancers ( $p < 0.005$ ). Detection of the hTERT protein by IHC has permitted further analysis of carcinogenesis and cancer diagnosis (34).

In recent years, there has been disagreement over the suitability of telomerase activity as a prognostic biologic marker in breast cancer that may help to differentiate patients for individually based risk-related therapy. Hiyama *et al* (15), in a study of 140 breast cancer specimens with the TRAP assay, found a strong association between telomerase activity and stage classification and observed telomerase activity in 68% of stage I tumors and 95% of stage IV tumors. Potemba *et al* (1), using tissue microarrays, found a statistically significant correlation between tumor-specific survival (overall survival) and hTERT expression in breast cancer. However, some problems in interpretation may affect this apparent consensus. First, some samples of breast cancer tissue may be extensively contaminated by infiltrating lymphocytes during operative manipulations, especially in advanced disease, causing overestimation of telomerase activity and/or hTERT expression. In our previous reports, higher expression levels

of signals for both hTERT mRNA and protein were detected in some sections containing both carcinoma and lymphocytes in thyroid and colorectal cancers (13,24). Secondly, Poremba *et al* (1) used polyclonal antibodies against hTERT protein as a signal for expression. In our hands, polyclonal antibodies against hTERT protein give rise to strong background signals and are not clearly specific for measuring expression in cancer tissues. We have carefully compared the reactivity against hTERT protein of the monoclonal antibody used in the present study with that of some polyclonal antibodies. Use of the monoclonal antibody in IHC allowed clear demonstration of hTERT protein expression, with results similar to those of ISH for hTERT mRNA expression. Furthermore, IHC is technically much easier to perform than ISH, since contamination of samples by RNase is not an issue in IHC.

To the best of our knowledge, this report is the first on the study of hTERT expression in breast cancer as a function of neo-adjuvant treatment. We examined hTERT expression in 4 cases of breast cancer before and after chemotherapy. Before chemotherapy, hTERT was strongly expressed in all 4 carcinomas, but after chemotherapy hTERT expression had completely disappeared in all 4 cases. hTERT expression by lymphocytes was detectable by ISH and IHC both before and after chemotherapy in all 4 cases, and the level of expression did not appear to be altered by treatment.

In conclusion, determination of hTERT mRNA expression by ISH and hTERT protein expression by IHC can be used to obtain information contributing to a histopathological diagnosis during screening of breast cancers. By use of a monoclonal antibody, we could very clearly and sensitively demonstrate hTERT protein expression in breast cancer tissues but not in non-cancerous tissues. We also demonstrated that 4 carcinomas with originally positive immunoreactivity against hTERT protein became negative after neo-adjuvant chemotherapy. These results suggest that determination of hTERT protein by IHC can be used as a potential diagnostic and therapeutic marker to distinguish breast cancers.

#### Acknowledgments

We thank Ms. Sachiko Nishimura for her expert technical assistance.

#### References

- Poremba C, Heine B, Diallo R, Heinecke A, Wai D, Schaefer KL, Braun Y, Schuck A, Lanvers C, Bankfalvi A, Kneif S, Thorst J, Zuber M, Kochli OR, Mross F, Dietrich H, Sauter G, Stein H, Fogt F and Boecker W: Telomerase as a prognostic marker in breast cancer: high-throughput tissue microarray analysis of hTERT and hTR. *J Pathol* 198: 181-189, 2002.
- Healy KC: Telomere dynamics and telomerase activation in tumor progression: prospects for prognosis and therapy. *Oncol Res* 7: 121-130, 1995.
- Shampay J and Blackburn EH: Tetrahymena micronuclear sequences that function as telomeres in yeast. *Nucleic Acids Res* 17: 3247-3260, 1989.
- Levy MZ, Allsopp RC, Futcher AB, Greider CW and Harley CB: Telomere end-replication problem and cell aging. *J Mol Biol* 225: 951-960, 1992.
- Moyzis RK, Buckingham JM, Cram LS, Dani M, Deaven LL, Jones MD, Meyne J, Ratliff RL and Wu JR: A highly conserved repetitive DNA sequence, (TTAGGG)<sub>n</sub>, present at the telomeres of human chromosomes. *Proc Natl Acad Sci USA* 85: 6622-6626, 1988.
- Kim NW, Piatyszek MA, Prowse KR, Harley CB, West MD, Ho PL, Coviello GM, Wright WE, Weinrich SL and Shay JW: Specific association of human telomerase activity with immortal cells and cancer. *Science* 266: 2011-2015, 1994.
- Blackburn EH: Telomerase. *Annu Rev Biochem* 61: 113-129, 1992.
- Shay JW and Bacchetti S: A survey of telomerase activity in human cancer. *Eur J Cancer* 33: 787-791, 1997.
- Meyerson M, Counter CM, Eaton EN, Ellisen LW, Steiner P, Caddle SD, Ziaugra L, Beijersbergen RL, Davidoff MJ, Liu Q, Bacchetti S, Haber DA and Weinberg RA: hEST2, the putative human telomerase catalytic subunit gene, is up-regulated in tumor cells and during immortalization. *Cell* 90: 785-795, 1997.
- Nakamura TM, Morin GB, Chapman KB, Weinrich SL, Andrews WH, Lingner J, Harley CB and Cech TR: Telomerase catalytic subunit homologs from fission yeast and human. *Science* 277: 955-959, 1997.
- Kilian A, Bowtell DD, Abud HE, Hime GR, Venter DJ, Keese PK, Duncan EL, Reddel RR and Jefferson RA: Isolation of a candidate human telomerase catalytic subunit gene, which reveals complex splicing patterns in different cell types. *Hum Mol Genet* 6: 2011-2019, 1997.
- Kamori M, Nakamura KI, Kanauchi H, Obara T, Kawahara M, Mimura Y, Kaminishi M and Takubo K: Consistent decrease in telomere length in parathyroid tumors but alteration in telomerase activity limited to malignancies: a preliminary report. *World J Surg* 26: 1083-1087, 2002.
- Kamori M, Nakamura K, Hashimoto M, Ogawa T, Kaminishi M and Takubo K: Clinical application of human telomerase reverse transcriptase gene expression in thyroid follicular tumors by fine-needle aspirations using *in situ* hybridization. *Int J Oncol* 22: 985-991, 2003.
- Poremba C, Scheel C, Hero B, Christiansen H, Schaefer KL, Nakayama J, Berthold F, Jurgens H, Boecker W and Dockhorn-Dworniczak B: Telomerase activity and telomerase subunit gene expression patterns in neuroblastoma: a molecular and immunohistochemical study establishing prognostic tools for fresh-frozen and paraffin-embedded tissues. *J Clin Oncol* 18: 2582-2592, 2000.
- Hiyama E, Gollahon L, Kataoka T, Kuroi K, Yokoyama T, Gazdar AF, Hiyama K, Piatyszek MA and Shay JW: Telomerase activity in human breast tumors. *J Natl Cancer Inst* 88: 116-122, 1996.
- Sugino T, Yoshida K, Bolodoku J, Tahara H, Bulcy J, Manek S, Wells C, Goodison S, Ide T, Suzuki T, Tahara E and Tarin D: Telomerase activity in human breast cancer and benign breast lesions: diagnostic applications in clinical specimens, including fine needle aspirates. *Int J Cancer* 69: 301-306, 1996.
- Nawaz S, Hashizumi TL, Markham NE, Shroyer AL and Shroyer KR: Telomerase expression in human breast cancer with and without lymph node metastases. *Am J Clin Pathol* 107: 542-547, 1997.
- Clark GM, Osborne CK, Levitt D, Wu F and Kim NW: Telomerase activity and survival of patients with node-positive breast cancer. *J Natl Cancer Inst* 89: 1874-1881, 1997.
- Poremba C, Boecker W, Willenbring H, Schaefer KL, Otterbach F, Burger H, Diallo R and Dockhorn-Dworniczak B: Telomerase activity in human proliferative breast lesions. *Int J Oncol* 12: 641-648, 1998.
- Hoos A, Hepp HH, Kaul S, Ahlert T, Basten G and Wallwiener D: Telomerase activity correlates with tumor aggressiveness and reflects therapy effect in breast cancer. *Int J Cancer* 79: 8-12, 1998.
- Roos G, Nilsson P, Cajander S, Nielsen NH, Arnerlov C and Landberg G: Telomerase activity in relation to p53 status and clinico-pathological parameters in breast cancer. *Int J Cancer* 79: 343-348, 1998.
- Mokbel K, Parris CN, Ghilchik M, Williams G and Newbold RF: The association between telomerase, histopathological parameters, and Ki-67 expression in breast cancer. *Am J Surg* 178: 69-72, 1999.
- Umbrecht CB, Sherman ME, Dome J, Carey LA, Marks J, Kim N and Sukumar S: Telomerase activity in ductal carcinoma *in situ* and invasive breast cancer. *Oncogene* 18: 3407-3414, 1999.
- Kamori M, Kanauchi H, Nakamura KI, Kawahara M, Weber TK, Mafune K, Kaminishi M and Takubo K: Demonstration of human telomerase reverse transcriptase (hTERT) in human colorectal carcinomas by *in situ* hybridization. *Int J Oncol* 20: 15-21, 2002.

25. Kamatori M, Nakamura KI, Ogawa T, Mafune KI, Tatsumi Y, Obara T, Onoda N, Fujiwara M, Izumiyama-Shimomura N, Mori M, Kaminishi M and Takubo K: Demonstration of human telomerase reverse transcriptase (hTERT) in human parathyroid tumors by *in situ* hybridization with a new oligonucleotide probe. *Clin Endocrinol* 58: 43-48, 2003.
26. Fukushima M, Shimomura N, Nakamura KI, Kamatori M, Koizumi K, Shimizu K and Takubo K: Demonstration of human telomerase reverse transcriptase by *in situ* hybridization in lung carcinoma. *Oncol Rep* 12: 1227-1232, 2004.
27. Frost M, Bobak JB, Gianani R, Kim N, Weinrich S, Spalding DC, Cass LG, Thompson LC, Enomoto T, Uribe-Lopez D and Shroyer KR: Localization of telomerase hTERT protein and hTR in benign mucosa, dysplasia, and squamous cell carcinoma of the cervix. *Am J Clin Pathol* 114: 726-734, 2000.
28. Kawakami Y, Kitamoto M, Nakanishi T, Yasui W, Tahara E, Nakayama J, Ishikawa F, Tahara H, Ide T and Kajiyama G: Immunohistochemical detection of human telomerase reverse transcriptase in human liver tissues. *Oncogene* 19: 3888-3893, 2000.
29. Pearson WR and Lipman DJ: Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* 85: 2444-2448, 1988.
30. Caruthers MH, Beaucage SL, Escavitch JW, Fisher EF, Goldman RA, De Haseth PL, Mandeck W, Matteucci MD, Rosendahl MS and Siabinsky Y: Chemical synthesis and biological studies on mutated gene-control regions. *Cold Spring Harbor Symp Quant Biol* 47: 411-418, 1982.
31. Tani Y: PCR *In situ* amplification and catalyzed signal amplification: approaches of higher sensitive, non-radioactive *in situ* hybridization. *Acta Histochem Cytochem* 32: 261-270, 1999.
32. Harvey JM, Clark GM, Osborne CK and Allred DC: Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. *J Clin Oncol* 17: 1474-1481, 1999.
33. Allred DC, Harvey JM, Berardo M and Clark GM: Prognostic and predictive factors in breast cancer by immunohistochemical analysis. *Mod Pathol* 11: 155-168, 1998.
34. Tahara H, Yasui W, Tahara E, Fujimoto J, Iio K, Tamai K, Nakayama J, Ishikawa F, Tahara E and Ide T: Immunohistochemical detection of human telomerase catalytic component, hTERT, in human colorectal tumor and non-tumor tissue sections. *Oncogene* 18: 1561-1567, 1999.



## Expression of the ubiquitin-proteasome pathway and muscle loss in experimental cancer cachexia.

Khal J, Wyke SM, Russell ST, Hine AV, Tisdale MJ.

Pharmaceutical Sciences Research Institute, Aston University, Birmingham, UK.

Muscle protein degradation is thought to play a major role in muscle atrophy in cancer cachexia. To investigate the importance of the ubiquitin-proteasome pathway, which has been suggested to be the main degradative pathway mediating progressive protein loss in cachexia, the expression of mRNA for proteasome subunits C2 and C5 as well as the ubiquitin-conjugating enzyme, E2(14k), has been determined in gastrocnemius and pectoral muscles of mice bearing the MAC16 adenocarcinoma, using competitive quantitative reverse transcriptase polymerase chain reaction. Protein levels of proteasome subunits and E2(14k) were determined by immunoblotting, to ensure changes in mRNA were reflected in changes in protein expression. Muscle weights correlated linearly with weight loss during the course of the study. There was a good correlation between expression of C2 and E2(14k) mRNA and protein levels in gastrocnemius muscle with increases of 6-8-fold for C2 and two-fold for E2(14k) between 12 and 20% weight loss, followed by a decrease in expression at weight losses of 25-27%, although loss of muscle protein continued. In contrast, expression of C5 mRNA only increased two-fold and was elevated similarly at all weight losses between 7.5 and 27%. Both proteasome functional activity, and proteasome-specific tyrosine release as a measure of total protein degradation was also maximal at 18-20% weight loss and decreased at higher weight loss. Proteasome expression in pectoral muscle followed a different pattern with increases in C2 and C5 and E2(14k) mRNA only being seen at weight losses above 17%, although muscle loss increased progressively with increasing weight loss. These results suggest that activation of the ubiquitin-proteasome pathway plays a major role in protein loss in gastrocnemius muscle, up to 20% weight loss, but that other factors such as depression in protein synthesis may play a more important role at higher weight loss.

PMID: 16160695 [PubMed - in process]



# Increased expression of proteasome subunits in skeletal muscle of cancer patients with weight loss.

Khal J, Hine AV, Fearon KC, Dejong CH, Tisdale MJ.

Pharmaceutical Sciences Research Institute, Aston University, Birmingham B4 7ET, UK.

Atrophy of skeletal muscle is common in patients with cancer and results in increased morbidity and mortality. In order to design effective therapy the mechanism by which this occurs needs to be elucidated. Most studies suggest that the ubiquitin-proteasome proteolytic pathway is most important in intracellular proteolysis, although there have been no reports on the activity of this pathway in patients with different extents of weight loss. In this report the expression of the ubiquitin-proteasome pathway in rectus abdominis muscle has been determined in cancer patients with weight loss of 0-34% using a competitive reverse transcriptase polymerase chain reaction to measure expression of mRNA for proteasome subunits C2 and C5, while protein expression has been determined by western blotting. Overall, both C2 and C5 gene expression was increased by about three-fold in skeletal muscle of cachectic cancer patients (average weight loss 14.5 $\pm$ 2.5%), compared with that in patients without weight loss, with or without cancer. The level of gene expression was dependent on the amount of weight loss, increasing maximally for both proteasome subunits in patients with weight loss of 12-19%. Further increases in weight loss reduced expression of mRNA for both proteasome subunits, although it was still elevated in comparison with patients with no weight loss. There was no evidence for an increase in expression at weight losses less than 10%. There was a good correlation between expression of proteasome 20S $\alpha$  subunits, detected by western blotting, and C2 and C5 mRNA, showing that increased gene expression resulted in increased protein synthesis. Expression of the ubiquitin conjugating enzyme, E2(14k), with weight loss followed a similar pattern to that of proteasome subunits. These results suggest variations in the expression of key components of the ubiquitin-proteasome pathway with weight loss of cancer patients, and suggest that another mechanism of protein degradation must be operative for patients with weight loss less than 10%.

PMID: 16125116 [PubMed - in process]





**Cell type-specific occurrence of caveolin-1alpha and -1beta in the lung caused by expression of distinct mRNAs.**

Kogo H, Aiba T, Fujimoto T.

Department of Anatomy and Molecular Cell Biology, Nagoya University Graduate School of Medicine, Showa-ku, Nagoya 466-8550, Japan. [hkogo@fujita-hu.ac.jp](mailto:hkogo@fujita-hu.ac.jp)

Two isoforms of caveolin-1, alpha and beta, had been thought to be generated by alternative translation initiation of an mRNA (FL mRNA), but we showed previously that a variant mRNA (5'V mRNA) encodes the beta isoform specifically. In the present study, we demonstrated strong correlation between the expression of the caveolin-1 protein isoforms and mRNA variants in culture cells and the developing mouse lung. The alpha isoform protein and FL mRNA were expressed constantly during the lung development, whereas expression of the beta isoform protein and 5'V mRNA was negligible in the fetal lung before 17.5 days post coitum, and markedly increased simultaneously at 18.5 days post coitum, when the alveolar type I cells started to differentiate. Immunohistochemical analysis revealed the cell type-specific expression of the two isoforms; the alveolar type I cell expresses the beta isoform predominantly, while the endothelium harbors the alpha isoform chiefly. The mutually exclusive expression of caveolin-1 isoforms was verified by Western blotting of the selective plasma membrane preparation obtained from the endothelial and alveolar epithelial cells. The present result indicates that the two caveolin-1 isoforms are generated from distinct mRNAs in vivo and that their production is regulated independently at the transcriptional level. The result also suggests that the alpha and beta isoforms of caveolin-1 may have unique physiological functions.

PMID: 15067006 [PubMed - indexed for MEDLINE]

**Oncogene and growth factor expression in ovarian cancer.**

Kommos F, Bauknecht T, Birmelin G, Kohler M, Tesch H, Pfeleiderer A.

Department of Gynaecology, Albert-Ludwig University, Freiburg, Germany.

The varying tumor-biological behavior of ovarian carcinomas probably influences both their operability and response to chemotherapy, which are the most relevant prognostic factors. The phenotype of different ovarian carcinomas is obviously associated with an activation of the EGF/TGF-alpha signal pathway, including c-myc and c-jun expression. Analysis of EGF-R, TGF-alpha, c-myc and c-jun expression in 33 stage III/IV, and 2 stage I/II ovarian carcinomas with biochemical, molecular-chemical and immunohistochemical methods showed a correlation between the mRNA and protein levels of EGF-R and TGF-alpha for tumors with low or high expressing rates. However, the concentration of measurable free EGF-Rs seems to depend on the amount of TGF-alpha expression by the tumors. The EGF-R binding ligand TGF-alpha is produced by epithelial tumor cells; stromal cells are usually TGF-alpha-negative, as shown by immunohistochemistry. High expression rates of EGF-R, TGF-alpha and c-myc were detected in 6, 7, and 10 out of 35 ovarian carcinomas, respectively. C-jun mRNA was detected in 18/19 cases studied. Non-malignant tissues originating from myometrium or ovary expressed no (or only small amounts of) EGF-R or TGF-alpha mRNA, whereas a high c-myc expression was found in 1/7 normal myometria, and in 2/5 normal ovaries. There was no strong correlation between EGF-R/TGF-alpha and c-myc/c-jun expression.(ABSTRACT TRUNCATED AT 250 WORDS)

PMID: 1502888 [PubMed - indexed for MEDLINE]



**Somatostatin receptors in primary human breast cancer: quantitative analysis of mRNA for subtypes 1-5 and correlation with receptor protein expression and tumor pathology.**

**Kumar U, Grigorakis SI, Watt HL, Sasi R, Snell L, Watson P, Chaudhari S.**

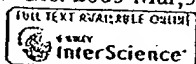
Fraser Laboratories For Diabetes Research, Department of Medicine, McGill University, Royal Victoria Hospital, 687 Pine Avenue West, H3A 1A1 Montreal, Quebec, Canada. [ujendra.kumar@muhc.mcgill.ca](mailto:ujendra.kumar@muhc.mcgill.ca)

Somatostatin receptors (SSTRs) have been identified in most hormone-producing tumors as well as in breast cancer. In the present study, we determined SSTR1-5 expression in primary ductal NOS breast tumors through semi-quantitative RT-PCR and immunocytochemistry. The results from the analysis of 98 samples were correlated with several key histological markers and receptor expression. All five SSTR subtypes are variably expressed at the mRNA level in breast tumors with 91% of samples showing SSTR1, 98% SSTR2, 96% SSTR3, 76% SSTR4, and 54% SSTR5. SSTR1-5 are localized to both tumor cells and the surrounding peritumoral regions as detected by immunocytochemistry. Levels of SSTR mRNA, when corrected for beta-actin levels, were highest for SSTR3 followed by SSTR1, SSTR2, SSTR5, and SSTR4. Furthermore, there was good correlation between mRNA and protein expression with 84% for SSTR1, 79% for SSTR2, 89% for SSTR3, 68% for SSTR4, 68% for SSTR5, and 78% for all five receptors. SSTR1, 2 and 4 were correlated with ER levels whereas SSTR2 showed an additional correlation with PR levels. These correlations were independent of patient age and histological grade. Moreover, using immunocytochemistry, blood vessels exhibited receptor-specific localization for SSTR2 and SSTR5. Our results indicate significant correlations between mRNA and protein expression along with receptor-specific correlations with histological markers as well as ER and PR levels. Differential distribution of SSTR subtypes in tumors and receptor-specific expression in vascular structures may be considered as a novel diagnosis for breast tumors with receptor subtype agonists.

**Publication Types:**

- Evaluation Studies

PMID: 15986128 [PubMed - indexed for MEDLINE]



## A transcriptomic and proteomic analysis of the effect of CpG-ODN on human THP-1 monocytic leukemia cells.

Kuo CC, Kuo CW, Liang CM, Liang SM.

Institute of BioAgricultural Sciences, Academia Sinica, Taipei, Taiwan.

The CpG motif of bacterial DNA (CpG-DNA) is a potent immunostimulating agent whose mechanism of action is not yet clear. Here, we used both DNA microarray and proteomic approaches to investigate the effects of oligodeoxynucleotides containing the CpG motif (CpG-ODN) on gene transcription and protein expression profiles of CpG-ODN responsive THP-1 cells. Microarray analysis revealed that 2 h stimulation with CpG-ODN up-regulated 50 genes and down-regulated five genes. These genes were identified as being associated with inflammation, antimicrobial defense, transcriptional regulation, signal transduction, tumor progression, cell differentiation, proteolysis and metabolism. Longer stimulation (8 h) with CpG-ODN enhanced transcriptional expression of 58 genes. Among these 58 genes, none except one, namely WNT1 inducible signaling pathway protein 2, was the same as those induced after 2 h stimulation. Proteomic analysis by two-dimensional gel electrophoresis, followed by mass spectrometry identified several proteins up-regulated by CpG-ODN. These proteins included heat shock proteins, modulators of inflammation, metabolic proteins and energy pathway proteins. Comparison of microarray and proteomic expression profiles showed poor correlation. Use of more reliable and sensitive analyses, such as reverse transcriptase polymerase chain reaction, Western blotting and functional assays, on several genes and proteins, nonetheless, confirmed that there is indeed good correlation between mRNA and protein expression after CpG-ODN treatment. This study also revealed that several anti-apoptotic and neuroprotective related proteins, not previously reported, are activated by CpG-DNA. These findings have extended our knowledge on the activation of cells by CpG-DNA and may contribute to further understanding of mechanisms that link innate immunity with acquired immune response(s).

PMID: 15693060 [PubMed - indexed for MEDLINE]

## REGULAR ARTICLE

This material may be protected by Copyright law (Title 17 U.S. Code)

# A transcriptomic and proteomic analysis of the effect of CpG-ODN on human THP-1 monocytic leukemia cells

*Cheng-Chin Kuo<sup>1</sup>, Chu-Wei Kuo<sup>2</sup>, Chi-Ming Liang<sup>3</sup> and Shu-Mei Liang<sup>1</sup>*<sup>1</sup> Institute of BioAgricultural Sciences, Academia Sinica<sup>2</sup> Graduate Institutes of Life Sciences, National Defense Medical Center<sup>3</sup> Institute of Biological Chemistry, Academia Sinica, Taipei, Taiwan, Republic of China

The CpG motif of bacterial DNA (CpG-DNA) is a potent immunostimulating agent whose mechanism of action is not yet clear. Here, we used both DNA microarray and proteomic approaches to investigate the effects of oligodeoxynucleotides containing the CpG motif (CpG-ODN) on gene transcription and protein expression profiles of CpG-ODN responsive THP-1 cells. Microarray analysis revealed that 2 h stimulation with CpG-ODN up-regulated 50 genes and down-regulated five genes. These genes were identified as being associated with inflammation, antimicrobial defense, transcriptional regulation, signal transduction, tumor progression, cell differentiation, proteolysis and metabolism. Longer stimulation (8 h) with CpG-ODN enhanced transcriptional expression of 58 genes. Among these 58 genes, none except one, namely WNT1 inducible signaling pathway protein 2, was the same as those induced after 2 h stimulation. Proteomic analysis by two-dimensional gel electrophoresis, followed by mass spectrometry identified several proteins up-regulated by CpG-ODN. These proteins included heat shock proteins, modulators of inflammation, metabolic proteins and energy pathway proteins. Comparison of microarray and proteomic expression profiles showed poor correlation. Use of more reliable and sensitive analyses, such as reverse transcriptase polymerase chain reaction, Western blotting and functional assays, on several genes and proteins, nonetheless, confirmed that there is indeed good correlation between mRNA and protein expression after CpG-ODN treatment. This study also revealed that several anti-apoptotic and neuroprotective related proteins, not previously reported, are activated by CpG-DNA. These findings have extended our knowledge on the activation of cells by CpG-DNA and may contribute to further understanding of mechanisms that link innate immunity with acquired immune response(s).

Received: May 13, 2004  
Revised: November 13, 2004  
Accepted: November 15, 2004

**Keywords:**

ADP-ribosylation factor 3 / CpG-ODN / Heat shock protein / Microarray

Correspondence: Dr. Shu-Mei Liang, Institute of BioAgricultural Sciences, Academia Sinica, Taipei 115, Taiwan, Republic of China  
E-mail: smyang@gate.sinica.edu.tw  
Fax: +886-2-26515120

Abbreviations: ARF3, ADP-ribosylation factor 3; FPKL1, formyl peptide receptor-like 1; HEK293, human embryonic kidney 293 cells; HSP, heat shock protein; IL, interleukin; JNK, c-Jun NH<sub>2</sub>-terminal kinase; LPS, lipopolysaccharide; LxR, nuclear receptor subfamily 1; MyD88, myeloid differentiation factor 88; PKC, protein kinase C gamma; PGK, phosphoglycerate kinase; TLR9, toll-like receptor 9; WISP-2, WNT1 inducible signaling pathway protein

## 1 Introduction

Mammals protect themselves against pathogen infection primarily via innate and adaptive immunity [1]. The innate immune system relies on a set of pattern recognition receptors (e.g., Toll-like receptors) to recognize foreign molecular structures such as lipopolysaccharide (LPS) and bacterial DNA [2, 3]. Innate immune cells recognize these molecular structures and initiate not only innate but also adaptive immunity by producing immunomodulatory cytokines and activating T and B immune cells [1]. Bacterial DNA can directly activate B cells to

proliferate and secrete immunoglobulins in a T cell-independent manner [4–6]. It also induces B cells and monocytes to activate transcription factor NF- $\kappa$ B and secrete cytokines, including interleukin (IL) 12, tumor necrosis factor  $\alpha$  (TNF- $\alpha$ ), and interferon  $\alpha/\beta$  [7–10]. The immunostimulatory activity of bacterial DNA has been assigned to unmethylated CpG motifs (GACGTT for murine, GTCGTT for human) [11]. Recent evidence shows that synthetic oligodeoxynucleotides containing a CpG motif (CpG-ODN), like bacterial DNA with the CpG moiety (CpG-DNA), induce potent Th1-like immune responses that are protective against several infectious agents and immune disorders in animal models [12, 13]. Biologically active CpG-ODN, like bacteria DNA, activates macrophages and immature dendritic cells to increase expression of MHC class II and costimulatory molecules, thereby transcribing cytokine mRNAs, and producing pro-inflammatory cytokines including TNF $\alpha$ , IL-1, IL-6 and IL-12 [9, 14–16]. CpG-ODN can therefore serve as an adjuvant and immunomodulator in vaccines against a wide variety of targets, including infectious agents, cancer antigens and allergens [17].

It has been suggested that unmethylated CpG-DNA-mediated immune activation functions through a toll-like receptor 9 (TLR9) signaling pathway [18]. Endocytosis and sequentially endosomal maturation as well as binding of heat shock protein (HSP) 90 to CpG-DNA are essential for induction of TLR9 signal transduction [19, 20]. It has also been shown that recognition of CpG-DNA causes TLR to form a dimer, which recruits the adaptor molecule, myeloid differentiation factor 88 (MyD88), through interaction between their C-terminal Toll/IL-1R domains. This recruitment of MyD88 to the Toll/IL-1R domain of TLR9 initiates a signaling pathway that sequentially involves IL-1R-associated kinase 1 and TNF- $\alpha$  receptor-associated factor 6 [18, 21, 22]. Studies using gene-deficient mice and RAW264.7 cells transiently transfected with dominant-negative forms of these molecules have indicated that the MyD88-mediated signaling pathway is essential for CpG-DNA-induced activation of NF- $\kappa$ B and c-Jun NH $_2$ -terminal kinase (JNK), as well as subsequent production of cytokines in monocytic cells [18, 21, 22]. The precise mechanism of action of CpG-DNA and CpG-ODN, nonetheless, is still not thoroughly understood. To further elucidate the molecular events after binding of CpG-ODN to TLR9, in this study, we treated CpG-ODN responsive THP-1 cells with CpG-ODN and evaluated changes by using DNA microarray and proteomic approaches. We have discovered up-regulation of more than 50 distinguished genes/proteins and identified induction of several anti-apoptotic and neuroprotecting genes by CpG-ODN treatment.

## 2 Materials and methods

### 2.1 Reagents

Phosphorothioate-modified CpG-ODN and CpG-ODN were synthesized by MDBio (Taipei, Taiwan). Human specific ODN sequences are: CpG-ODN, 5'-TCC TCG TTT TGT CGT

TTT GTC GTT-3'; CpC-ODN, 5'-TGC TGC TTT TGT GCT TTT GTG CTT-3'. The mouse specific CpG-ODN sequence is 5'-TCC ATG ACG TTC CTG ATG CT-3'. CHCA was from Sigma (St. Louis, MO, USA).

### 2.2 Cell culture

Cell lines were obtained from the American Type Culture Collection (Rockville, MD). Mouse RAW264.7 macrophage and human embryonic kidney 293 cells (HEK293) were cultured in DMEM supplemented with 10% heat inactivated fetal bovine serum, 100 U/mL penicillin, 100  $\mu$ g/mL streptomycin sulfate, 200 mmol/L L-glutamine, and 50  $\mu$ M  $\beta$ -mercaptoethanol in a humidified atmosphere of 5% CO $_2$  at 37°C. The medium was changed every 2 days for all experiments. Human THP-1 monocytic leukemia cells, which have been shown to express TLR9 and respond to CpG-DNA stimulation [23, 24], were cultured in RPMI1640 with the same supplements as for RAW264.7 cell cultures.

### 2.3 Human cDNA microarray

Total RNAs extracted from cultured THP-1 cells were isolated with TRIzol (Invitrogen, Leek, The Netherlands) and submitted to Genasia Biotechnology (Taipei, Taiwan) for further processing. In brief, 4  $\mu$ g of total RNA from CpG-ODN stimulated, or normal THP-1 cells was labeled with a fluorescence marker (U-vision, Taipei, Taiwan). Different colored fluorescence dyes (Cy5 and Cy3) were used to distinguish total RNA from normal and ODN stimulated cells. The labeled RNA was used for hybridization with the Human 1 cDNA microchip from Agilent Technologies (Palo Alto, CA, USA). The chips were scanned and the expression pattern was analyzed using genechip software. Genes showing up-regulation or down-regulation of RNA levels were analyzed and identified on a genomic database as suggested by the manufacturer of the microchip.

### 2.4 Protein preparation

THP-1 cells were seeded in a 175 cm $^2$  tissue culture flask at a density of 10 $^6$  cells per milliliter in culture medium. The cells were stimulated with or without 1.5  $\mu$ M CpG-ODN at defined times and harvested by centrifugation at 4°C, 1000  $\times$  g for 15 min. Cell pellets were washed twice with ice-cold PBS, resuspended and sonicated in extraction buffer containing 25 mM Tris-HCl (pH 7.5), 2 mM  $\beta$ -mercaptoethanol and protease inhibitor cocktail. After centrifugation at 10 000  $\times$  g for 20 min, ammonium sulfate was added to the supernatant until the final concentration reached 50% saturation w/v. The solution was stirred at 4°C for 30 min and centrifuged at 10 000  $\times$  g for 30 min at 4°C. The supernatant fraction was then transferred into a fresh tube, and the precipitated protein pellet solubilized in extraction buffer. To remove salts and other contaminants, the extracts were treated with a pre-cooled (–20°C) solution of 10% TCA in acetone with 0.07%  $\beta$ -mercaptoethanol. Proteins were allowed to precipitate overnight at –20°C.

After centrifugation, the pellet was washed with ice-cold acetone, containing 0.07%  $\beta$ -mercaptoethanol. The supernatant was discarded and the pellet dried in a SpeedVac system (Model AES1010; Savant, Holbrook, NY, USA).

## 2.5 2-DE

2-DE was performed using an IPGphor IEF and a Hofer DALT vertical unit (Amersham Biosciences, Piscataway, NJ, USA). One milligram of dried protein sample was dissolved in 350  $\mu$ L of rehydration buffer solution, containing 7 M urea, 2 M thiourea, 4% w/v CHAPS, 5 mM tributyl phosphine, and 2% IPG and loaded onto an immobilized pH 3–10 linear gradient strip (18 cm), followed by rehydration for 16 h. IEF was then performed in the following manner: 100 V for 30 min, 250 V for 30 min, 500 V for 30 min, 1000 V for 30 min, 4000 V for 30 min, 6000 V for 55 000 Vh. At the end of IEF, the IPG strips were equilibrated for 15 min in buffer containing 6 M urea, 2% w/v SDS, 30% v/v glycerol, and 50 mM Tris, pH 6.8, then reduced with 65 mM dithioerythritol (DTE) and subsequently alkylated with 135 mM iodoacetamide for another 15 min. After equilibration, the IPG strips were immediately placed on top of a 12% SDS-PAGE (1.5 mm, 20  $\times$  24 cm). The second dimension gels were then overlaid with molten 0.8% agarose solution in SDS electrophoresis buffer. Electrophoresis was performed at 16°C, starting at 10 mA per gel for 1 h, followed by 45 mA per gel until the dye front reached the bottom of the gels.

## 2.6 Staining and image acquisition

Immediately after electrophoresis, gels were stained with SYPRO Ruby (Molecular Probes, Eugene, OR, USA). In brief, gels were fixed for 30 min in 10% methanol, 7% acetic acid, and then stained overnight in SYPRO Ruby stain. The staining solution was removed and gels were washed in 10% methanol and 7% acetic acid for 3 h. After staining, image acquisition was carried out on a Typhoon 9200 (Amersham Biosciences). To identify a protein, spot detection, quantification and matching of 2-D results were analyzed using ImageMaster software (Amersham Biosciences). The  $M_r$  of the proteins were calibrated according to the LMW-SDS Marker Kit (Amersham Biosciences), and their  $pI$  values were estimated from the position of the protein spots on the 2-D gel and confirmed with the information supplied by the manufacturer. Since most of the  $pI$  values for the truncated proteins had not been reported previously, the  $pI$  values of the truncated proteins were estimated from the position of the observed spots. To omit the variation due to the use of separate gels, after background subtraction, the intensity levels of protein spots on each gel were normalized as a proportion of one reference spot, and protein quantities were calculated by integrating the density over the spot area. Protein spots that showed reproducible modulation exceeding  $\sim$ 80% after CpG-ODN treatment in three experiments were further analyzed by MS.

## 2.7 In-gel digestion with trypsin and extraction of peptides

Protein spots were excised from stained gels and cut into pieces. In brief, gel spots were dehydrated with ACN for 10 min and dried in a vacuum centrifuge. Gel pieces were reswelled with 55 mM DTE in 25 mM ammonium bicarbonate (pH 8.5) at 37°C for 1 h. The solution was then exchanged with alkylation solution, which contained 100 mM iodoacetamide in 25 mM ammonium bicarbonate (pH 8.5), at room temperature for 1 h. After alkylation, the gel pieces were washed twice with 50% ACN in 25 mM ammonium bicarbonate (pH 8.5) for 15 min. The wash solution was discarded and the pieces of gel were dehydrated with ACN for 10 min and dried in a vacuum centrifuge. Trypsin digestion was initiated by reswelling the gel in 25 mM ammonium bicarbonate solution with 25 ng of trypsin (Promega, Madison, WI, USA). After incubation at 37°C for 16 h, tryptic peptides were extracted twice with 50% ACN containing 5% formic acid for 15 min with moderate sonication. The extracted solutions were pooled and evaporated to dryness in a vacuum centrifuge. The dried peptide mixture was dissolved in 0.1% formic acid and used for MS.

## 2.8 MALDI-Q-TOF MS and protein identification

Tryptic peptides analyses were performed using a Micromass Q-TOF Ultima MALDI (Micromass, Wythenshawe, U.K.) equipped with a 337 nm nitrogen laser and operated in reflectron positive ion mode. Peptide mixtures (1  $\mu$ L) were premixed with 1  $\mu$ L of the matrix (5 mg CHCA in 50% ACN with 0.1% TFA) then spotted onto the MALDI target plate. Mass spectra were acquired for the mass range of 900–3500 Da and the individual spectra from MALDI MS or MS/MS were processed using the Micromass Masslynx 4.0 software. The generated peak list files were used to query the Swiss-Prot database using the MASCOT program (<http://www.matrixscience.com>) with the following parameters: peptide mass tolerance, 50 ppm; MS/MS ion mass tolerance, 0.25 Da; allowance of missed cleavage, 1; and consideration for variable modifications such as oxidation of methionine and carboxyamidomethylation of cysteines. Only significant hits as defined by MASCOT probability analysis were considered initially. In addition, when the PMF matches were between 5 and 9, at least one peptide sequence was manually checked by MALDI MS/MS analysis.

## 2.9 RT-PCR analysis

cDNA from THP-1 cells was produced with Superscript II reverse transcriptase (Invitrogen) using a oligo(dT)<sub>18</sub> primer for 1 h at 42°C. PCR of cDNA was performed using specific primers for the gene of interest and control  $\beta$ -actin. All PCR products were electrophoresed on a 1.5% agarose gel, and DNA bands were visualized by staining the gel with ethidium bromide.

## 2.10 Immunoblotting

Human THP-1 or mouse macrophage RAW264.7 cells ( $5 \times 3 \times 10^6$ /well) were cultured in a six-well culture plate and treated with or without  $1.5 \mu\text{M}$  CpG-ODN for the designated times. After stimulation, cells were harvested by centrifugation at  $1000 \times g$  for 15 min in a refrigerated centrifuge and washed twice with cold PBS buffer. The cells were lysed on ice for 15 min with 300  $\mu\text{L}$  lysis buffer (Pierce, Rockford, USA), supplemented with protease inhibitor cocktail (Sigma). The lysates were centrifuged at  $12000 \times g$  for 15 min at  $4^\circ\text{C}$ , and protein concentrations of supernatant were determined using the Bio-Rad Protein Assay (Hercules, CA, USA). The lysates (50  $\mu\text{g}$  of protein/lane) were subjected to 12% SDS-PAGE and transferred to NC membranes (Amersham Biosciences). The membranes were blocked in PBS-0.1% Tween 20 (PBST) containing 5% non-fat skim milk at room temperature for 1 h, followed by staining with anti-ADP-ribosylation factor 3 (ARF-3) monoclonal antibody (0.1  $\mu\text{g}/\text{mL}$ ; Sigma). The membranes were then incubated with horseradish peroxidase-conjugated secondary antibody (dilution, 1:3000) for 1 h. After washing three times with PBST, specific bands were detected by chemiluminescence according to the manufacturer's protocol (Amersham Biosciences).

## 2.11 Cell transfection and luciferase assay

HEK293 cells ( $5 \times 3 \times 10^6$ /well) were transfected using FuGENE 6 (Roche Molecular Biochemicals, Indianapolis, IN, USA) plus 0.1  $\mu\text{g}$  p5xNF- $\kappa\text{B}$ -luc (Stratagene, La Jolla, CA, USA), 0.05  $\mu\text{g}$  pCDNA3.1- $\beta$ -galactosidase, and pCDNA3.1-hTLR9 overnight. The cells were incubated with or without  $1.5 \mu\text{M}$  CpG-ODN for 8 h and then lysed. NF- $\kappa\text{B}$  luciferase activity assays were performed as recommended by the manufacturer (Promega).  $\beta$ -galactosidase activity was used to normalize the data.

## 2.12 Enzyme activity assay

Pyruvate kinase activity was assayed in a solution (1 mL) containing 100 mM Tris-HCl (pH 8.0), 100 mM KCl, 10 mM  $\text{MgCl}_2$ , 0.2 mM NADH, 10 mM PEP, 1.5 mM ADP, 1 unit of lactate dehydrogenase, and an appropriate amount of cell lysate from CpG-ODN untreated or treated THP-1 cells. The reaction was monitored at  $30^\circ\text{C}$  for a period of time by measuring the decrease in absorbance at 340 nm. PKG activity was assayed in a coupled reaction with glyceraldehyde 3-phosphate dehydrogenase (GAPDH) as described by Lee [25]. In brief, the assay was performed at  $30^\circ\text{C}$  in a total volume of 1 mL containing 100 mM Tris-HCl (pH 7.9), 10 mM  $\text{MgCl}_2$ , 0.15 mM NADH, 2 mM ATP, 6 mM 3-phosphoglycerate, 0.1 mg/mL BSA, 50 mg of GAPDH, and an appropriate amount of cell lysate. NADH consumption was monitored at 340 nm.

## 3 Results

### 3.1 Effect of CpG-ODN on gene expression profiles of human THP-1 cells

To elucidate the effect of CpG-ODN on gene expression, THP-1, a cell line known to express TLR9 and respond to CpG-DNA [23, 24] was cultured with or without CpG-ODN. Since preliminary experiments showed that  $1.5 \mu\text{M}$  CpG-ODN caused more contrasting results between normal and CpG-ODN treated cells,  $1.5 \mu\text{M}$  CpG-ODN was used throughout the experiments. To distinguish CpG-ODN treated samples from the control, total cellular RNA of normal and CpG-ODN treated cells was isolated and labeled with the fluorescence dyes, Cy5 and Cy3, respectively. The labeled RNA was then used for hybridization with a Human 1 cDNA microchip from Agilent Technologies. Of the 13000 human genes represented on the gene array, a total of 55 genes changed expression significantly after 2 h CpG-ODN treatment. Among these, 50 genes were up-regulated while five genes were down-regulated by a factor  $\geq 2$ . These genes were sorted by functions and are listed in Tables 1 and 2. They included notably, IL-18 receptor accessory protein, MSGA beta gene, thioredoxin, pro-pol-dUTPase polypeptide, Sp140, connexin 59 gene, Grb2-like 2, enoyl-coenzyme A hydratase, propionyl coenzyme A carboxylase, cytochrome P450, and WNT1 inducible signaling pathway protein 2 (WISP-2) etc. The function of these genes are known to be related to inflammatory responses, antimicrobial defense, transcriptional regulation, intracellular signal transduction, tumor progression, cell differentiation, proteolysis etc.

Table 1. Genes up-regulated ( $\geq 2$  fold) in human THP-1 cells after 2 h of CpG-ODN treatment

Gene name and description (changed fold $\geq 2$ )	Genebank number
<b>Inflammation and receptor</b>	
IL-18 receptor accessory protein	AF077346
T cell receptor V beta gene	X58806
MSGA, beta gene	U03019
Antigen gene (PA)	M21896
TIED	NM_004791
Platelet activating receptor	AF002986
<b>Antimicrobial defence</b>	
Thioredoxin	NM_003329
Pro-Pol-dUTPase polypeptide	AC004748
Nuclear body protein Sp140	U63420
<b>Transcriptional regulation</b>	
Putative transcription factor LUZP	A1986271
General transcription factor II, I, pseudogene 1	A1700706
Connexin 59 gene	L29277
Basic helix-loop-helix protein class B 1 (BHLHB1)	AF221520



Table 1. Continued

Gene name and description (changed fold $\geq 2$ )	Genebank number
<b>Regulatory protein</b>	
Advillin	AF041449
<b>Channel and transport</b>	
Small GTP binding protein Rab9	U44103
FXD domain-containing ion transport regulator 1	AI125364
<b>Signal transduction</b>	
SH3 domain Grb2-like 2	AF036268
Titin	X90568
Titin associated protein (165 kD protein)	X69089
KIAA1451 protein	AB040884
Vasoactive intestinal peptide receptor	U11087
<b>Enzyme and protease</b>	
Enoyl-Coenzyme A hydratase	AI800553
Nephrin, B-type metallopeptidase	U65090
Propionyl Coenzyme A carboxylase	AB011145
Acyloxyacyl hydrolase (neutrophil)	M62840
Cytochrome P450	U79716
Intestinal alkaline phosphatase	M31008
<b>Tumor progression and cell differentiation</b>	
Retinoblastoma 1	L11910
Human genomic DNA of 9q32 anti-oncogene of flat Epithelium cancer, segment 6/10	AB036268
WNT1 inducible signaling pathway protein 2 (WISP-2)	AF100780
<b>Structure protein</b>	
Collagen, type IV, alpha 6	D21337
Beta Myosin heavy chain	M58018
<b>Other</b>	
Homo sapiens Cri-du-chat region mRNA, clone NIBB11	U52827
Human mRNA for laminin alpha 5 chain, partial cds.	AB010099
NIK like and Thyroxin-binding globulin precursor	Z83850
Hypothetical protein DKFZp434M0331	AL137720
Hypothetical protein FLJ11021 similar to splicing factor	AK023985
Hypothetical protein	AL049851
Chromosome 18 open reading frame 1	NM_004338
Arfaptin 1	AW408785
Zinc finger protein 8 (ZFP8)	M29581
Zinc finger protein 137 (clone pHZ-30)	U09414
Olfactomedin related ER localized protein	AI738468
Cyclin-dependent kinase 8	BE467537
Integrin, alpha 1	D87462
KIAA0421 protein	AB007881
KIAA1233 protein	AB033059
Unnamed protein product	AK026362
NBL4	X75535
BC331191_1	AAD39268

Table 2. Genes down-regulated ( $\geq 2$  fold) in human THP-1 cells after 2 h of CpG-ODN treatment

Gene name and description (changed fold $\geq 2$ )	Genebank number
Zinc-finger homeodomain protein 4	BAB03600
Human protein kinase MEKK2b mRNA, complete cds.	AF239798
Glypican 5	U66033
Human genomic DNA, chromosome 22q11.2, clone N75A12.	AP000362
Collagen, type I, alpha 1	Z74615

Longer stimulation of THP-1 cells with CpG-ODN (8 h) resulted in the up-regulation of 58 genes. These genes included notably IL-10 receptor beta, formyl peptide receptor-like 1 (FPR1), vitamin D receptor, nuclear receptor subfamily 1 (LxR), early B-cell factor, protein kinase C gamma (PKC), Nck, Ash, phospholipase C binding protein (NAP4), phosphoriboxyl pyrophosphate amidotransferase, disheveled 3, WISP-2 etc. Analysis of the functions of the 58 up-regulated genes showed that they are associated with anti-inflammation, transcriptional regulation, intracellular signal transduction, tumor progression, cell differentiation, proteolysis, neurodegeneration, neuroprotection etc. (Table 3). We also found that the stimulation of THP-1 cells with CpG-ODN for different periods of times resulted in different profiles. Several defense related genes such as IL-18 receptor accessory protein, Pro-Pol-dUTPase polypeptide, Sp140 and connexin 59 were transiently up-regulated at 2 h short stim-

Table 3. Genes up-regulated ( $\geq 2$  fold) in human THP-1 cells after 8 h of CpG-ODN treatment

Gene name and description (changed fold $\geq 2$ )	Genebank number
<b>Inflammation and receptor</b>	
IL-10 receptor (beta)	U08988
Formyl peptide receptor-like 1 (FPR1)	AF081535
Vitamin D receptor	J03258
NMDAR1	Z32774
CD44 antigen	AW028346
Nuclear receptor subfamily 1 (LxR)	NM_005693
Neuromedin B receptor	M73482
<b>Transcriptional regulation</b>	
Early B-cell factor (ebf)	AF208502
Neurogenic differentiation 1 (Neuro D)	AB018693
MAX dimerization protein (NESH protein)	AB037886
Ribosomal protein S6 kinase	AF090421
ASH2L	AB022785
<b>Regulatory protein</b>	
Hypothetical protein DKFZp434H0820	AL137555
Peroxisomal farnesylated protein	X75535
LTBP4	AF051344
Neuronal pentraxin II	U29195

Table 3. Continued

Gene name and description (changed fold $\geq 2$ )	Genebank number
<b>Channel and transport</b>	
Gamma-aminobutyric [61] A receptor	NM_004961
ATP synthase subunit F6	M37104
Transient receptor potential channel 1	Z73903
Choroideremia (Rab escort protein 1)	X57637
<b>Signal transduction</b>	
Protein kinase C, gamma	Z15114
Regulator of G-protein signaling 5	AI674877
Nck, Ash and phospholipase C binding protein (NAP4)	AB005216
Highly similar to adenylyl kinase gene	AB016886
<b>Enzyme and protease</b>	
Phenylalanine hydroxylase	AA203389
Carboxypeptidase A1	X67318
Xylulokinase	AK001205
Pancreatic lipase	J05125
Ubiquitin specific protease 12	AF022789
Transmembrane protease, serine2	U75329
Aspartate beta-hydroxylase	U03109
Phosphoribosyl pyrophosphate amidotransferase	D13757
<b>Tumor progression and cell differentiation</b>	
CDC23	AF053977
WISP-2	AF100780
Microseminoprotein, beta	M34376
Dishevelled 3	NM_004423
<b>Structure protein</b>	
Trichohyalin	L09190
Kertarin	AF061809
<b>Other</b>	
Human transferrin pseudogene	M22376
TIMP-2	U44383
Collagen-like protein	U67921
Human genomic DNA, chromosome 21q, section 60/105	AP001716
Human genomic DNA, chromosome 21q, section 64/105	AP001720
KIAA0136	D50926
KIAA0379	AB002377
KIAA0489	AB007958
KIAA1114	AL049732
KIAA1451	AB040884
KIAA0756	AB018299
Zinc finger protein 267	AF220492
Hypothetical protein FLJ10633	AK001495
Hypothetical protein EUOIMAGE 1955967	AK026108
Myb1 homolog like 1	AK001893
Antizyme inhibitor	D88674
Disintegrin-like and metalloprotease (reprolysin type) with Thrombospondin type 1 motif, 3	AB002364
ADP-ribosylation factor 3 (ARF-3)	M74491
Testis specific protein, Y-linked	M98525
Unnamed protein	AK026042

Table 4. List of antimicrobial and anti-inflammatory genes modulated by CpG-ODN treatment of THP-1 cells

Gene name	Genebank number	Expression fold	
		2 h	8 h
Connexin 59 gene	L29277	2.12 $\pm$ 0.05	1.74 $\pm$ 0.08
IL-18 receptor accessory protein	X58806	2.32 $\pm$ 0.21	1.19 $\pm$ 0.13
Integrin, alpha 1	X68742	2.01 $\pm$ 0.03	1.05 $\pm$ 0.10
Nuclear body protein Sp140	U63420	2.22 $\pm$ 0.11	1.39 $\pm$ 0.04
Pro-Pol-dUTPase polyprotein	AC004748	2.33 $\pm$ 0.18	0.95 $\pm$ 0.03
Thioredoxin	NM_003329	2.20 $\pm$ 0.08	1.07 $\pm$ 0.01
FPRL1	AF081535	0.95 $\pm$ 0.06	2.13 $\pm$ 0.11
IL-10 receptor	U08988	1.22 $\pm$ 0.18	2.21 $\pm$ 0.07
LxR	NM_005693	0.90 $\pm$ 0.03	2.37 $\pm$ 0.31
Vitamin D receptor	J03258	1.37 $\pm$ 0.23	2.39 $\pm$ 0.11

Expression fold is designated as the ratio of CpG-ODN treated over control

ulation but were down-regulated thereafter, while anti-inflammatory associated genes such as FPRL1, IL10 receptor, vitamin D receptor and LxR were up-regulated after 8 h stimulation (Tables 1, 3 and 4).

### 3.2 Verification of the microarray results with RT-PCR or Western blotting

To verify the results from the microarray analysis, we also performed RT-PCR on the up-regulated genes (Fig. 1). Consistent with results obtained in the microarray gene expression analysis, RT-PCR studies showed that the mRNA levels of some selected genes, including ubiquitin specific protease 12, regulator of G-protein signaling 5, NAP4 and ASH2L, were increased in response to CpG-ODN (Fig. 1, Table 3). In addition, the protein expression level of ARF-3

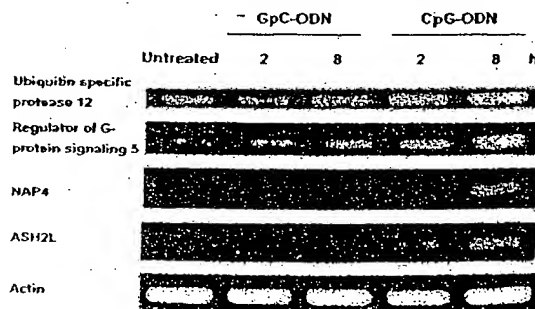
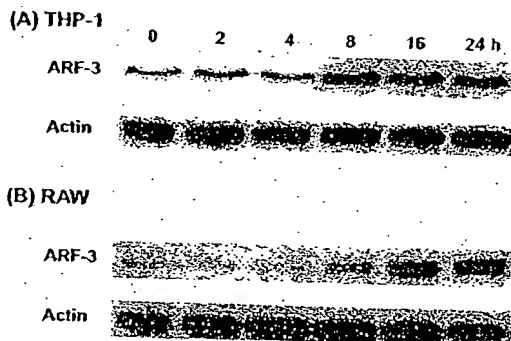


Figure 1. Induction of various genes by CpG-ODN. THP1 cells were stimulated with medium alone, 1.5  $\mu$ M GpC-ODN (as the negative control) or CpG-ODN for the indicated times. RT-PCR was then performed to analyze gene expression levels.  $\beta$ -actin was used as an internal control. The experiment was repeated three times with similar results.



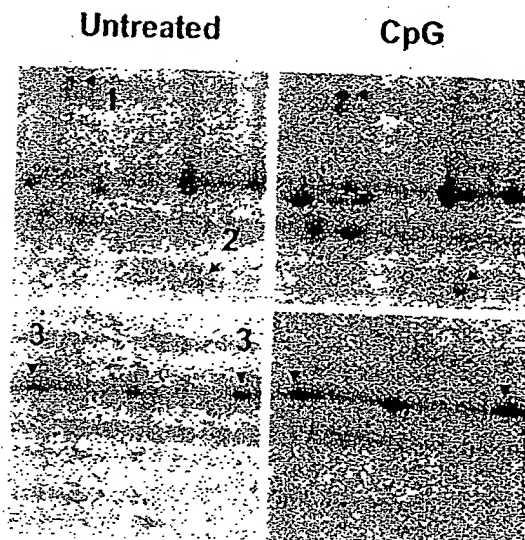
**Figure 2.** CpG-ODN induced ARF-3 protein expression in human THP-1 or mouse RAW264.7 cells. THP-1 (A) or mouse RAW264.7 (B) cells were incubated with 1.5  $\mu$ M CpG-ODN for the indicated time points. The protein expression level of ARF-3 was determined by Western blotting of cell extracts using anti-ARF-3 antibody. The experiment was repeated three times with similar results.

was shown to increase in Western blotting analysis in cell lysates from THP-1 cells treated with CpG-ODN for 8–24 h (Fig. 2A). Similar studies showed that the ARF-3 protein was also induced by mouse specific CpG-ODN in other TLR9 expression cell lines such as the mouse macrophage RAW264.7 cell line (Fig. 2B).

### 3.3 Proteins regulated in CpG-ODN stimulated THP-1 cells

To further assess whether there was any correlation between regulation of gene expression and expression of cellular proteins, a proteomic approach was adopted to identify protein expression profiles. THP-1 cells were treated with CpG-ODN for defined times (from 8 to 40 h), and their cytoplasmic proteins were extracted for 2-DE analysis. Although the use of high concentrations of urea might give us a broader view of all the proteins affected by CpG-ODN, preliminary results from 2-D gels showed that the resolution of the protein mixtures were not satisfactory. To improve and get the best resolution from 2-DE, total proteins were roughly separated into supernatant and precipitated fractions using 50% saturated ammonium sulfate solution. To remove salts and other contaminants, both protein fractions were precipitated with TCA solution and then subjected to 2-DE. By protein spot determination analysis, about 500 and 450 well-resolved spots were observed on each pH 3.0–10.0 gel for precipitated or supernatant fractions, respectively. Comparative analysis of 2-DE between treatments and control showed that the intensities of the protein spots from the ammonium sulfate precipitated fraction did not change, while several protein spots were up-regulated by at least ~80% in the supernatant fraction of 8 h CpG-ODN stimulated THP-1 cells (Fig. 3).

The protein spots were individually excised from gels for further identification. After trypsin digestion, several protein spots were identified without ambiguity by MS MALDI-



**Figure 3.** The effect of CpG-ODN on the 2-DE profile of THP-1 cells. THP-1 cells were treated with or without 1.5  $\mu$ M CpG-ODN for 8 h. Total proteins were extracted and roughly separated into two fractions by 50% saturation with ammonium sulfate. The supernatant fraction was then separated by 2-DE. Protein spots were visualized by SYPRO Ruby staining. Comparison of CpG-ODN treated THP-1 cells to untreated cells showed that these proteins changed in intensity by over 80%. Protein spots were identified by trypsin digestion and MS. Localization of protein spots 1 (enoyl-coenzyme A hydratase), 2 (proteasome  $\alpha$ ) and 3 (cyclophilin A; two isoforms) are shown. The experiment was repeated three times with similar results.

TOF. These proteins included HSP60, HSP90, cyclophilin A, enoyl-coenzyme A hydratase, eukaryotic translation elongation factor, proteasome  $\alpha$  and  $\beta$  chain and ATP synthase beta chain (Table 5). Similar experiments on cells treated for a longer period of time with CpG-ODN stimulation (25 h) revealed that 27 protein spots were changed in intensity by at least ~80%. These protein spots contained members of HSPs (HSP27, hsc70, grp78 and grp94), metabolic enzymes (phosphoglycerate kinase (PGK) and pyruvate kinase (PYK)), macrophage capping protein and cyclophilin A (Table 6). Among these proteins, macrophage capping protein, PGK, PYK, cyclophilin A and HSP27 (Figs. 4 and 5A) were found to be up-regulated. Interestingly, we found that a truncated form of grp78 with an expected mass of 25 kDa and pI of 5.3 was up-regulated while grp78 itself was down-regulated. A similar situation was also found for grp94 and hsc70 and their truncated derivatives (Table 7 and Fig. 5). In addition, we also observed six down-regulated protein spots on 2-D gels in samples after 25 h CpG-ODN treatment. Among these six proteins, we have successfully identified three as 40s ribosomal protein SA, grp78 and hsc70, respectively (Table 6), while the other three, due to their relative low abundance, have not been identified yet.

Table 5. List of proteins modulated by 8 h CpG-ODN treatment

Protein name	Accession no.	$M_r$ (theor.)	pI (theor.)	Matched no.	Coverage%	Score	Expression fold
ATP synthase beta-chain	gi114549	56 525	5.26	16	58	171	2.33 $\pm$ 0.06
Cyclophilin A	P05092	17 870	7.82	5	35	62	2.85 $\pm$ 0.13
Enoyl-Coenzyme A hydratase	gi4503447 <sup>a)</sup>	35 971	6.61	11	44	62	2.52 $\pm$ 0.05
Eukaryotic translation elongation factor	gi4503481	50 087	6.25	6	25	68	3.41 $\pm$ 0.21
HSP60	P10809 <sup>b)</sup>	57 963	5.24	13	27	76	2.78 $\pm$ 0.03
HSP90-beta	P08238	83 133	4.97	10	18	65	2.36 $\pm$ 0.10
Proteasome $\alpha$ chain	gi4506181	25 882	6.92	11	59	80	2.52 $\pm$ 0.11
Proteasome $\beta$ chain	gi4506193	26 472	8.27	9	46	84	3.85 $\pm$ 0.17

Expression fold is designated as the ratio of CpG-ODN treated over control

a) NCBI accession number

b) Swiss-Prot accession number

Table 6. List of proteins modulated by 16 and 25 h CpG-ODN treatment

Protein name	Swiss-Prot no.	$M_r$ (theor.)	pI (theor.)	Matched no.	Coverage%	Score	Expression fold	
							16h	25h
Cyclophilin A	P05092	17 870	7.82	5	35	62	2.53 $\pm$ 0.02	2.48 $\pm$ 0.15
78 kDa glucose regulated protein (grp78)	P11021	72 288	5.07	13	30	148	0.61 $\pm$ 0.04	0.29 $\pm$ 0.06
HSP27	P04792	22 768	5.98	12	59	124	1.00 $\pm$ 0.01	2.61 $\pm$ 0.12
Heat shock cognate 70 kDa protein (hsc70)	P11142	70 854	5.37	16	34	114	0.64 $\pm$ 0.01	0.31 $\pm$ 0.03
Macrophage capping protein	P40121	38 494	5.88	9	30	58	1.31 $\pm$ 0.01	2.58 $\pm$ 0.03
Phosphoglycerate kinase	P00558	44 284	7.052	11	33	71	2.32 $\pm$ 0.11	4.23 $\pm$ 0.19
Pyruvate kinase	P14618	57 710	7.95	17	32	114	1.65 $\pm$ 0.07	2.70 $\pm$ 0.12
40s ribosomal protein SA (RSP40)	P08865	32 833	4.79	5	23	61	0.33 $\pm$ 0.02	0.35 $\pm$ 0.05

Table 7. List of truncated proteins detected in THP-1 cells after 25 h CpG-ODN treatment

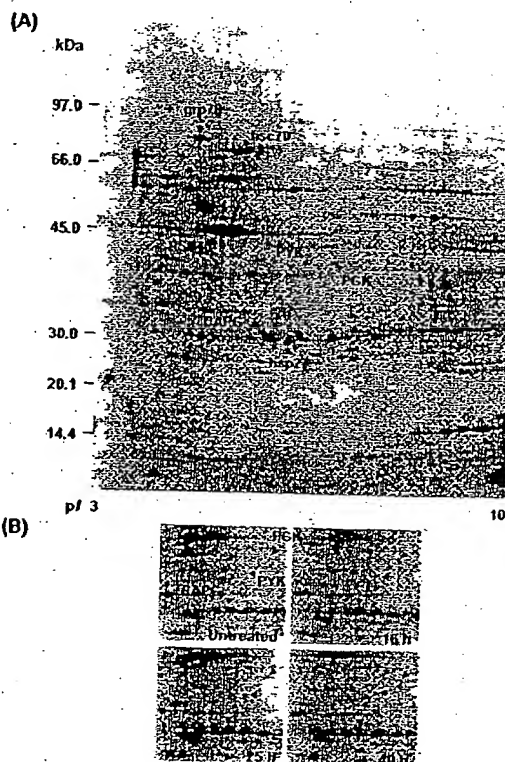
Protein name	Swiss-Prot no.	$M_r$ (obs.)	pI (obs.)	Matched no.	Coverage%	Score	Expression fold
94 kDa glucose-regulated protein (grp94)	P14625	~59 700	~5.00	12	14	104	New <sup>a)</sup>
Truncated form of grp78	P11021	~25 000	~5.30	12	22	75	New
Truncated form of hsc70	P11142	~22 000	~5.80	12	20	96	New
Truncated form of hsc70	P11142	~19 000	~6.10	11	18	113	New

a) New designated proteins detected in the CpG-ODN treated gel but not in the corresponding control gel

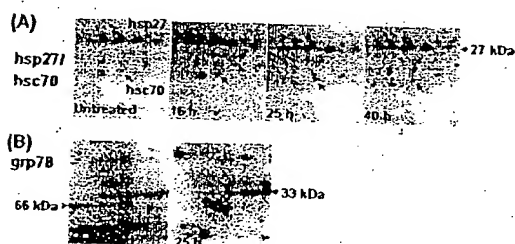
### 3.4 Comparison of microarray and proteomic results

Table 8 shows the expression of six genes and their corresponding proteins that were modulated by 8 h treatment of THP-1 cells with CpG-ODN. Besides enoyl-coenzyme A hydratase, there was poor correlation between the expression

of genes and their corresponding proteins (Table 8), suggesting that more in-depth studies were needed. To further evaluate whether changes observed in protein expression correlated with changes in mRNA levels, we randomly chose two proteins (PGK and PYK) that were induced after 16 h CpG-ODN treatment and determined their mRNA levels by



**Figure 4.** 2-D gel electrophoretic analysis of CpG-ODN-treated THP-1 cells. (A) Total cell protein from unstimulated THP-1 cells was subjected to 2-DE. (B) THP-1 cells were treated with or without 1.5 μM CpG-ODN for defined times. Cellular proteins were extracted and separated by 2-DE. Several up-regulated proteins are shown in the SYPRO Ruby staining gel. Comparison of CpG-ODN treated THP-1 cells to untreated cells showed that these proteins changed in intensity by over 80%. The experiment was repeated three times with similar results.

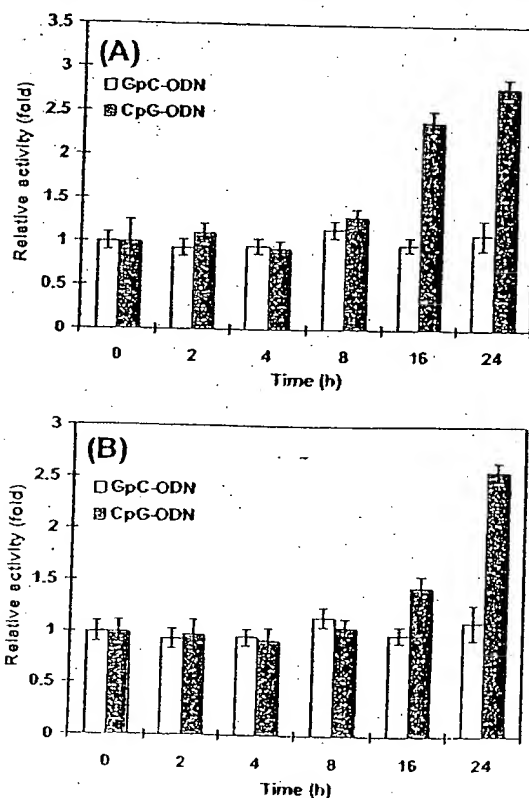


**Figure 5.** 2-D gel electrophoretic analysis of CpG-ODN-treated THP-1 cells. THP-1 cells were treated with or without 1.5 μM CpG-ODN for defined times. Cellular proteins were extracted and separated by 2-DE. Protein spots were detected by SYPRO Ruby staining. (A) Expression of HSP27 was induced by increasing the period of CpG-ODN stimulation. A truncated form of hsc70 was detected on the gel. (B) The native form of grp78 was detected in untreated cells, while the truncated form of grp78 was observed after 25 h stimulation. The experiment was repeated three times with similar results.

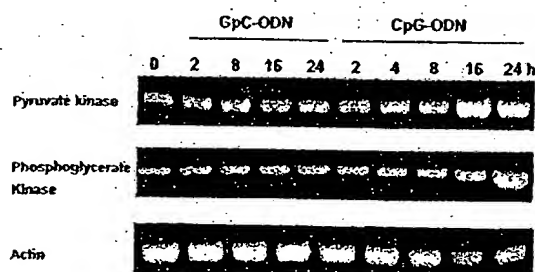
**Table 8.** Comparison of gene and protein expression levels in THP-1 cells after 8 h CpG-ODN treatment

Protein name	Gene expression fold from microarray	Protein expression fold from 2-D gel
Enoyl-Coenzyme A hydratase	1.72 ± 0.31	2.52 ± 0.05
Eukaryotic translation elongation factor	0.95 ± 0.01	3.41 ± 0.21
HSP60	0.94 ± 0.07	2.78 ± 0.03
HSP90-beta	1.58 ± 0.11	2.36 ± 0.10
Proteasome α chain	0.98 ± 0.03	2.52 ± 0.11
Proteasome β chain	1.09 ± 0.11	3.85 ± 0.17

**RT-PCR.** Our results showed that mRNA levels of PYK increased after 16 h CpG-ODN treatment, while mRNA levels of PGK were dramatically increased after 24 h stimulation (Fig. 6). In addition, we also performed enzyme activity analysis and found that the activity of PGK and PYK were



**Figure 6.** Activities of PYK and PGK induced by CpG-ODN. THP1 cells were stimulated with medium alone, 1.5 μM CpC-ODN (as the negative control) or CpG-ODN for the indicated times. Cell lysates were extracted and assayed for (A) PYK and (B) PGK activities. Data represent mean ± SEM. (n = 3).



**Figure 7.** Induction of PYK and PGK transcripts by CpG-ODN. THP1 cells were stimulated with medium alone, 1.5  $\mu$ M CpC-ODN or CpG-ODN for the indicated times. RT-PCR was then performed to analyze gene expression.  $\beta$ -actin was used as an internal control. The experiment was repeated three times with similar results.

indeed increased by a factor  $\sim 2.5$  after CpG-ODN stimulation (Fig. 7). To confirm that mRNA induced by CpG-ODN would also be accompanied by an increase in protein expression even though it was not detected in 2-D gel analysis, we used more sensitive and specific Western blotting analysis. As shown in Table 3 and Fig. 2, ARF-3 was identified in the microarray gene profile but not in the proteomic expression profile. Nevertheless, we observed enhanced protein expression of ARF-3 after CpG-ODN stimulation by Western blotting. Moreover, to investigate whether the up-regulation of ARF-3 by CpG-ODN is mediated through the TLR9 pathway, TLR9-deficient HEK293 cells were transiently cotransfected with hTLR9 and luciferase-reporter gene driven by a NF- $\kappa$ B-dependent promoter. Our data showed that NF- $\kappa$ B activity of untransfected HEK293 cells were not responsive to CpG-ODN stimulation, while in transfected HEK293 cells expressing hTLR9, NF- $\kappa$ B luciferase activity was up-regulated 12-fold after 8 h CpG-ODN stimulation. The activation of NF- $\kappa$ B induced by CpG-ODN was blocked by pretreatment of the transfected cells with an ARF-3 inhibitor, such as brefeldin A (Fig. 8), suggesting CpG-ODN induces ARF-3 and activates NF- $\kappa$ B after the interaction of CpG-ODN with TLR9.

#### 4 Discussion

In this study both microarray and proteomic approaches were used to evaluate the effect of CpG-ODN on gene/protein expression profiles of THP-1 cells at several time points. Comparison of the gene expression profiles showed that stimulation of the cells with CpG-ODN for different periods of time resulted in different profiles (Tables 1–4). The differences in mRNA expression between the cells with short and long stimulation could be attributable to the low reproducibility. However, to avoid experimental variations, we not only used the same batch of microarrays from the same manufacturer but also applied the samples of short and long term stimulation at the same time. In this way, we found that the

changes in expression fold of mRNA after CpG-ODN treatment were quite reproducible as shown by their mean  $\pm$  SEM (Table 4). A more likely explanation for the difference in the expression level of mRNA after different periods of stimulation with CpG-ODN is that the transient increase or decrease in these mRNA by CpG-ODN plays a significant role in modulating biological functions. For example, we found that the IL-18 receptor accessory protein from THP-1 cells was up-regulated after 2 h of CpG-ODN stimulation. The IL-18/IL18R system is known to activate Th1-mediated immune responses that play a critical role in host defense against infection [26]. Together with IL-18/IL18 R, several genes for antimicrobial defense were also increased, including thioredoxin, Pro-Pol-dUTPase polyprotein and Sp140. After 8 h of CpG-ODN stimulation, however, none of these genes was activated any more (Table 4). Since sustained or excessive production of these antimicrobial molecules might lead to inflammation and cellular damage [27], a plausible explanation is that THP-1 cells fight against the invasion of pathogens by up-regulating antimicrobial defense-associated genes at an early stage of stimulation and then shut them down to avoid over-activation. Whether this explanation is true remains to be verified.

It is noteworthy that our data also identified the up-regulation of several anti-inflammatory associated genes after 8 h of CpG-ODN stimulation. These genes included FPRL1, IL-10 receptor, vitamin D receptor, and LxR (Table 3). FPR and FPRL1 have been defined as chemotactic factors involved in host defense against bacterial infection and in the clearance of damaged cells. Additional studies have indicated that FPRL1 interacts with a menagerie of structurally diverse pro- and anti-inflammatory ligands associated with diseases, including amyloidosis, Alzheimer's diseases, prion disease and HIV [28, 29]. Therefore, FPRL1 may play an important role in regulating and/or balancing the production of pro- and anti-inflammatory molecules in CpG stimulated THP-1 cells. Additionally, a recent study has demonstrated that LxRs and their ligands act as negative regulators of macrophage inflammatory gene expression and inhibit the expression of inflammatory mediators such as inducible nitric oxide synthase, cyclooxygenase and IL-6 in response to bacterial infection or LPS stimulation [30]. Of interest, we found that a transcription factor gene connexin 59, a regulator of IL-6 expression, was up-regulated after 2 h of CpG-ODN stimulation. It is thus likely that CpG-ODN stimulation of THP-1 cells for 2 h may induce the expression of the pro-inflammatory cytokine IL-6 through the up-regulation of the connexin 59 gene, while 8 h of CpG-ODN treatment may counterbalance the initial inflammatory response by inducing LxR to inhibit IL-6 production.

Signal transduction molecules play an important role in cellular activation. Intracellular signal transduction systems employing various intermolecular interactions through docking elements, including SH2 and SH3 domains, have been reported [31–33]. Here we found that THP-1 cells treated with CpG-ODN for 2 h up-regulated gene expression

of Grb2-like protein (which contains an SH3 domain), while 8 h of stimulation induced Nck, Ash and phospholipase C binding protein (NAP4 which contains an SH2 domain). It is thus possible that Grb2-like protein and NAP4 may play important roles in CpG-ODN mediated signaling pathways. Furthermore, recent studies have also revealed that binding of CpG-DNA to TLR9 results in activation of JNK [34]. Since JNK is activated by Nck adaptor protein and Nck interacting kinase [35, 36], it is possible that CpG-ODN may activate JNK via up-regulation of NAP4. Although a recent publication described the gene expression profiles of a cultured mouse macrophage cell line after CpG-DNA stimulation [34], their microarray results were only conducted at one time-point (6 h stimulation). Moreover, they did not report the measurement of protein expression profiles in response to CpG-ODN stimulation.

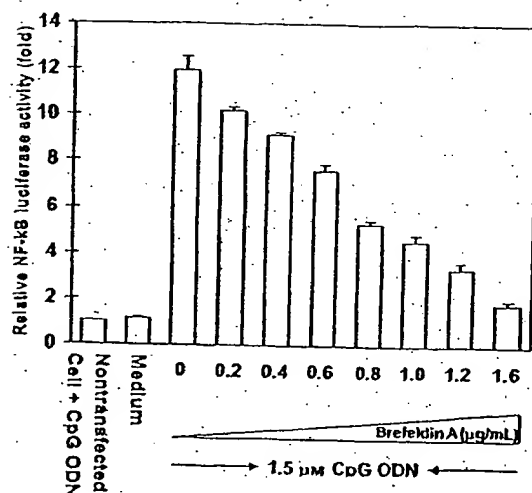
Comparison of the gene and protein expression profiles showed that there was discordance between mRNA and protein levels (Table 8). Similar discordance between the expression pattern of genes and proteins was also reported in other system using different stimuli [37–41]. The discordance between mRNA and protein levels could be due to screening capability such as detection sensitivity, choice of cut-off point, quantitativity of microarray and 2-D gels, as well as time discrepancy between gene and protein expression [39, 40, 42, 43]. Alternatively, it could also be explained by post-transcriptional events, such as alternative splicing or PTM [39, 40, 42, 43]. Another possible explanation is that most of the spots observed in the 2-D gels are isoforms of some proteins. The intensity of each spots does not necessarily represent total amount of a certain protein and thus does not correlate with its mRNA level. Our finding that microarray results correlated better with Western blotting results (e.g., ARF-3 in Fig. 2), an approach more suitable than 2-D gels for determining the total amount rather than isoforms of a given protein, seems to suggest that formation of isoforms should be carefully taken into consideration when one tries to correlate mRNA and protein expression data.

Using a proteomics approach, we found that CpG-ODN treatment up-regulated the expression of many proteins including HSPs, metabolic enzymes, structural proteins, as well as macrophage capping protein, cyclophilin and proteasome  $\alpha$  and  $\beta$  chain *etc.* HSPs are the most abundant and ubiquitous soluble intracellular proteins. They are up-regulated by various stressors including temperature, glucose deprivation, microbial infection and cancer [44]. They function as molecular chaperones to prevent protein aggregation and contribute to the folding of nascent and altered proteins. In addition, they are able to regulate immune responses, including production of inflammatory cytokines and chemokines and activation or maturation of immune cells [45, 46]. Beside HSPs, cyclophilin as well as proteasome  $\alpha$  and  $\beta$  chain have also been reported to be involved in the immune response [39, 47]; proteasome  $\beta$  chain is consistently up-regulated in human neutrophils following LPS exposure [39]. Our finding that the protein levels of HSPs, cyclophilin, and

proteasome  $\alpha$  and  $\beta$  chain were increased after CpG-ODN treatment suggests that these molecules might play a role in the immunostimulating effect of CpG-ODN. To what extent these proteins contribute to the immune responses of the cells to CpG-ODN is currently under study. Proteomic analysis also showed that truncated forms of grp78, grp94 and hsc70 were induced, a phenomenon similar to calreticulin observed by Richards and his coworkers [48]. The expression of full length hsc70 and grp78 were decreased while the levels of their truncated derivatives was increased after CpG-ODN treatment. These results suggest that the degradation of these proteins has been enhanced. We also found that proteasome  $\alpha$  and  $\beta$  chains as well as ubiquitin specific protease 12 were increased by CpG-ODN. Whether these enzymes or other enzymes were responsible for the generation of truncated hsc70 and grp78 remains to be elucidated.

Cells rely on multiple signaling pathways to determine their fates of survival, proliferation or apoptosis [49]. In fact, apoptosis plays an important role in regulating pathogen infection. To be able to grow and replicate in the target cells, pathogens may have to block apoptosis. Results from several laboratories have made it clear that HSP70 and HSP27 protect cells not only from heat, but also from most apoptotic stimuli [48, 50] by binding to Akt and subsequently mediating anti-apoptotic activity through activation of Akt [51–53]. Since our data revealed that CpG-ODN induced the expression of HSP90 and HSP27, it is possible that CpG-ODN might prevent apoptosis by up-regulation of HSPs.

Interestingly, our microarray data also showed that CpG-ODN mediated the induction of a set of genes associated with tumor progression and cell proliferation. Among these, one gene, WISP-2, was up-regulated by CpG-ODN after both 2 and 8 h stimulation. WISP genes were first identified as downstream targets of the Wnt-1 $\beta$ -catenin signaling pathway. They belong to the CCN family of growth factors that have been receiving increasing attention lately due to some of the family members having been reported to be involved in angiogenesis and tumorigenesis [54]. It would be interesting to evaluate whether CpG-ODN plays a role in angiogenesis and tumorigenesis by regulating WISP-2. In addition, we found that some genes associated with neurodegeneration or neuroprotection, such as FPRL1, NMDAR (NMDA) receptor, PKC and dishevelled 3 were up-regulated. To our knowledge, this is the first report to suggest an association between these genes and CpG-ODN stimulation. As mentioned above, FPRL1 plays a crucial role in proinflammatory aspects of systemic amyloidosis and neurodegenerative disease such as Alzheimer's disease and prion disease [28]. NMDAR, PKC and dishevelled are involved in modulating amyloid precursor protein metabolism, which is central to the pathogenesis of Alzheimer's disease [55–57]. Most notably, recent studies have shown that the TLR4-dependent pathway is involved in neurodegeneration of the central nervous system [58]. Whether CpG-ODN moieties of pathogens play any role in neurodegenerative diseases such as Alzheimer's remains to be elucidated.



**Figure 8.** ARF-3 participates in the CpG-ODN-TLR9-NF-κB pathway. HEK293 cells were cotransfected with p5xNF-κB and human TLR9. After overnight transfection, the cells were incubated with or without 1.5 μM CpG-ODN for 8 h in the presence or absence of increasing concentrations of the ARF-3 inhibitor brefeldin A. After incubation, cells were lysed and NF-κB luciferase activity was measured. Data represent mean ± SEM. ( $n = 3$ ).

Exposure of cells to LPS or microbial infection has been known to induce several genes encoding metabolic enzyme [34, 39]. Our microarray data also revealed that a large number of genes encoding proteins involved in energy synthesis and fatty acid oxidation, such as enoyl-coenzyme A hydratase, propionyl coenzyme A carboxylase and cytochrome p450 were activated by CpG-ODN treatment. In addition, we found that other proteins such as ARF-3 were up-regulated (Table 3, Fig. 8). ARFs are 20 kDa GTPases of the ras superfamily that are critical to vesicular trafficking, including exocytic protein transport and endocytosis [59, 60]. This study demonstrates for the first time that ARF-3 is involved in the activation of NF-κB induced by CpG-ODN (as shown in Fig. 8).

CpG-DNA/ODN has been shown to elicit primarily responses via the TLR9/MyD88 dependent pathway [18, 21, 22]. Chromosome location analysis showed that instead of localizing on one or two chromosome, the genes/proteins modulated by CpG-ODN stimulation are scattered on all chromosomes except chromosomes 23 and 24. These results seem to suggest that CpG-ODN either affects multiple chromosomes simultaneously or subsequently via cascades of cellular messengers. More studies are needed to elucidate its mechanism of actions.

## 5 Concluding remarks

In summary, by using microarray and proteomic approaches to evaluate the effect of CpG-ODN at different time points, we have found that genes/proteins regulated by CpG-ODN

are related to inflammatory responses, antimicrobial defense, transcriptional regulation, intracellular signal transduction, tumor progression, cell differentiation, proteolysis, anti-apoptosis as well as neurodegeneration and neuroprotection. Our results may help delineate the CpG-ODN mediated pathway and contribute to further understanding of mechanisms that link innate immunity with acquired immune response(s).

We thank Mr. Yen-Chieh Huang and Ms. V. R. Kavitha for technical assistance in RT-PCR analysis. We also thank the Core Facilities for Proteomics Research at the Academia Sinica, Taiwan for mass spectrometry analyses. This work was supported by the National Science Council (Grant NSC 91-3112-P001-002-Y) and Academia Sinica (Grant AS 911BC3PP), Republic of China.

## 6 References

- [1] Akira, S., Hemmi, H., *Immunol. Lett.* 2003, **85**, 85–95.
- [2] Janeway, C. A. Jr., Medzhitov, R., *Semin. Immunol.* 1998, **10**, 349–350.
- [3] Wagner, H., *Curr. Opin. Microbiol.* 2002, **5**, 62–69.
- [4] Halpern, M. D., Kurlander, R. J., Pisetsky, D. S., *Cell. Immunol.* 1996, **167**, 72–78.
- [5] Messina, J. P., Gilkeson, G. S., Pisetsky, D. S., *J. Immunol.* 1991, **147**, 1759–1764.
- [6] Yi, A. K., Krieg, A. M., *J. Immunol.* 1998, **160**, 1240–1245.
- [7] Tokunaga, T., Yamamoto, H., Shimada, S., Abe, H. et al., *J. Natl. Cancer Inst.* 1984, **72**, 955–962.
- [8] Yamamoto, S., Kuramoto, E., Shimada, S., Tokunaga, T., *Jpn. J. Cancer Res.* 1988, **79**, 866–873.
- [9] Stacey, K. J., Sweet, M. J., Hume, D. A., *J. Immunol.* 1996, **157**, 2116–2122.
- [10] Yi, A. K., Tuetken, R., Redford, T., Waldschmidt, M. et al., *J. Immunol.* 1998, **160**, 4755–4761.
- [11] Krieg, A. M., Yi, A. K., Matson, S., Waldschmidt, T. J. et al., *Nature* 1995, **374**, 546–549.
- [12] Kline, J. N., Waldschmidt, T. J., Businga, T. R., Lemish, J. E. et al., *J. Immunol.* 1998, **160**, 2555–2559.
- [13] Shirota, H., Sano, K., Hirasawa, N., Terui, T. et al., *J. Immunol.* 2001, **167**, 66–74.
- [14] Lipford, G. B., Sparwasser, T., Bauer, M., Zimmermann, S. et al., *Eur. J. Immunol.* 1997, **27**, 3420–3426.
- [15] Sparwasser, T., Miethke, T., Lipford, G., Borschert, K. et al., *Nature* 1997, **386**, 336–337.
- [16] Sparwasser, T., Miethke, T., Lipford, G., Erdmann, A. et al., *Eur. J. Immunol.* 1997, **27**, 1671–1679.
- [17] Takeda, K., Akira, S., *Genes Cells* 2001, **6**, 733–742.
- [18] Hemmi, H., Takeuchi, O., Kawai, T., Kaisho, T. et al., *Nature* 2000, **408**, 740–745.
- [19] Hacker, H., Mischak, H., Miethke, T., Liptay, S. et al., *EMBO J.* 1998, **17**, 6230–6240.
- [20] Bandholtz, L., Guo, Y., Palmberg, C., Mattsson, K. et al., *Cell. Mol. Life Sci.* 2003, **60**, 422–429.



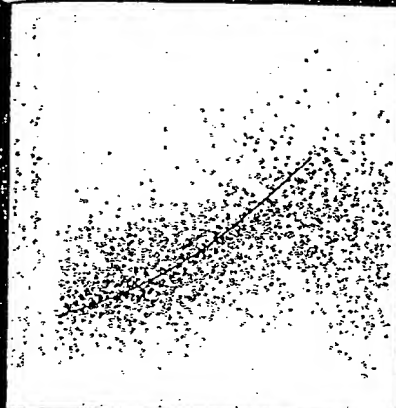
- [21] Bauer, S., Kirschning, C. J., Hacker, H., Redecke, V. et al., *Proc. Natl. Acad. Sci. USA* 2001, 98, 9237–9242.
- [22] Hacker, H., Vabulas, R. M., Takeuchi, O., Hoshino, K. et al., *J. Exp. Med.* 2000, 192, 595–600.
- [23] Takeshita, F., Leifer, C. A., Gursel, I., Ishii, K. J. et al., *J. Immunol.* 2001, 167, 3555–3558.
- [24] Akhtar, M., Watson, J. L., Nazli, A., McKay, D. M., *FASEB J.* 2003, 17, 1319–1321.
- [25] Lee, C. Y., *Methods Enzymol.* 1982, 90 Pt E, 121–126.
- [26] Kawakami, K., *J. Immunother.* 2002, 25 Suppl. 1, S12–19.
- [27] Nakanishi, K., Yoshimoto, T., Tsutsui, H., Okamura, H., *Annu. Rev. Immunol.* 2001, 19, 423–474.
- [28] Le, Y., Oppenheim, J. J., Wang, J. M., *Cytokine Growth Factor Rev.* 2001, 12, 91–105.
- [29] Le, Y., Murphy, P. M., Wang, J. M., *Trends Immunol.* 2002, 23, 541–548.
- [30] Joseph, S. B., Castrillo, A., Laffitte, B. A., Mangelsdorf, D. J., Tontonoz, P., *Nat. Med.* 2003, 9, 213–219.
- [31] Cohen, G. B., Ren, R., Baltimore, D., *Cell* 1995, 80, 237–248.
- [32] Pawson, T., *Nature* 1995, 373, 573–580.
- [33] Birge, R. B., Knudsen, B. S., Besser, D., Hanafusa, H., *Genes Cells* 1996, 1, 595–613.
- [34] Gao, J. J., Diesl, V., Wittmann, T., Morrison, D. C. et al., *J. Leukocyte Biol.* 2002, 72, 1234–1245.
- [35] Minden, A., Lin, A., Claret, F. X., Abo, A., Karin, M., *Cell* 1995, 81, 1147–1157.
- [36] Becker, E., Huynh-Do, U., Holland, S., Pawson, T. et al., *Mol. Cell Biol.* 2000, 20, 1537–1545.
- [37] Anderson, L., Seilhamer, J., *Electrophoresis* 1997, 18, 533–537.
- [38] Gygi, S. P., Rochon, Y., Franza, B. R., Aebersold, R., *Mol. Cell Biol.* 1999, 19, 1720–1730.
- [39] Fessler, M. B., Malcolm, K. C., Duncan, M. W., Worthen, G. S., *J. Biol. Chem.* 2002, 277, 31291–31302.
- [40] Kim, C. H., Kim do, K., Choi, S. J., Choi, K. H. et al., *Proteomics* 2003, 3, 2454–2471.
- [41] Scheurer, S. B., Raybak, J. N., Rosli, C., Neri, D., Elia, G., *Proteomics* 2004, 4, 1737–1760.
- [42] Corthals, G. L., Wasinger, V. C., Hochstrasser, D. F., Sanchez, J. C., *Electrophoresis* 2000, 21, 1104–1115.
- [43] Hegde, P. S., White, I. R., Debouck, C., *Curr. Opin. Biotechnol.* 2003, 14, 647–651.
- [44] Robert, J., *Dev. Comp. Immunol.* 2003, 27, 449–464.
- [45] Basu, S., Srivastava, P. K., *Cell Stress Chaperones* 2000, 5, 443–451.
- [46] Berwin, B., Nicchitta, C. V., *Traffic* 2001, 2, 690–697.
- [47] Fluckiger, S., Fijten, H., Whitley, P., Blasor, K., Cramer, R., *Eur. J. Immunol.* 2002, 32, 10–17.
- [48] Richards, J., Le Naour, F., Hanash, S., Beretta, L., *Ann. N. Y. Acad. Sci.* 2002, 975, 91–100.
- [49] Underhill, D. M., Ozinsky, A., *Annu. Rev. Immunol.* 2002, 20, 825–852.
- [50] Somersan, S., Larsson, M., Fonteneau, J. F., Basu, S. et al., *J. Immunol.* 2001, 167, 4844–4852.
- [51] Rane, M. J., Pan, Y., Singh, S., Powell, D. W. et al., *J. Biol. Chem.* 2003, 278, 27828–27835.
- [52] Solit, D. B., Basso, A. D., Olshen, A. B., Scher, H. I., Rosen, N., *Cancer Res.* 2003, 63, 2139–2144.
- [53] Sato, S., Fujita, N., Tsuruo, T., *Proc. Natl. Acad. Sci. USA* 2000, 97, 10832–10837.
- [54] Pennica, D., Swanson, T. A., Welsh, J. W., Roy, M. A. et al., *Proc. Natl. Acad. Sci. USA* 1998, 95, 14717–14722.
- [55] Lipton, S. A., *Cell Death Differ.* 1999, 6, 943–951.
- [56] Du, J., Zhou, S., Coggeshall, R. E., Carlton, S. M., *Neuroscience* 2003, 118, 547–562.
- [57] Mudher, A., Chapman, S., Richardson, J., Asuni, A. et al., *J. Neurosci.* 2001, 21, 4987–4995.
- [58] Lehnardt, S., Massillon, L., Follett, P., Jensen, F. E. et al., *Proc. Natl. Acad. Sci. USA* 2003.
- [59] Lenhard, J. M., Kahn, R. A., Stahl, P. D., *J. Biol. Chem.* 1992, 267, 13047–13052.
- [60] Morinaga, N., Adamik, R., Moss, J., Vaughan, M., *J. Biol. Chem.* 1999, 274, 17417–17423.
- [61] Balog, R. P., de Souza, Y. E., Tang, H. M., DeMasellis, G. M. et al., *Anal. Biochem.* 2002, 309, 301–310.

# PROTEOMICS

www.proteomics-journal.de

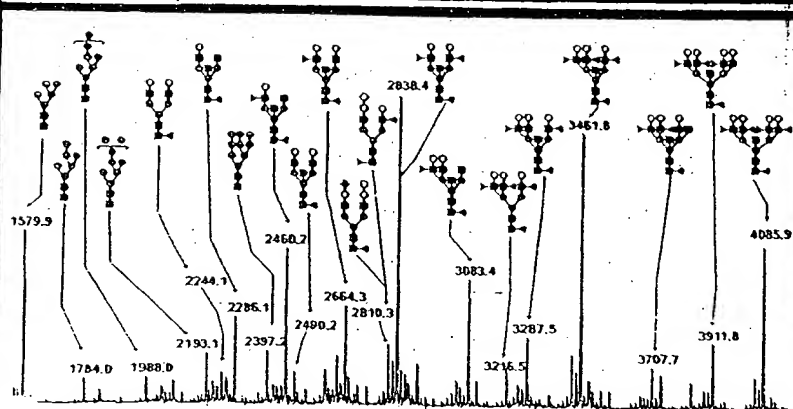
Q  
7P93847

4'05



## IPC '03

Proceedings of the 3<sup>rd</sup> International  
Proteomics Conference (IPC '03)  
held conjointly with the  
1<sup>st</sup> Taiwan Proteomics Conference and the  
2<sup>nd</sup> AOHUPO Congress  
Taipei, Taiwan, 14–17 May 2004

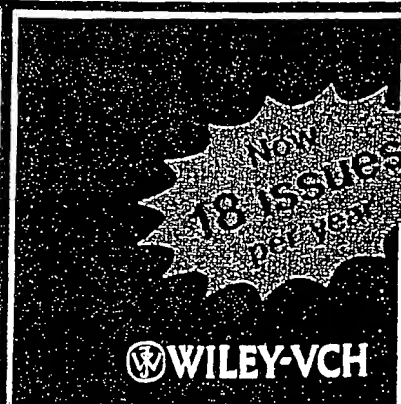


EBLING LIBRARY  
UNIVERSITY OF WISCONSIN

MAR 28 2005

750 Highland Avenue  
Madison, WI 53705

Editor:  
Richard J. Simpson



ISSN 1615-9853 • PROTC 5 (4) 831–1184 (2005) • Vol. 5 • No. 4 • March 2005

# CONTENTS

Volume 5- Issue 4  
March 2005  
Proteomics 5 (4) 831-1184 (2005)

## SPECIAL

### IPC'03

Proceedings of the 3<sup>rd</sup> International Proteomics  
Conference held conjointly with the  
1<sup>st</sup> Taiwan Proteomics Conference and the  
2<sup>nd</sup> AOHUPO Congress  
Taipei, Taiwan, 14-17 May 2004

Editor: Richard J. Simpson

- 831 EDITORIAL  
IPC '03  
Richard J. Simpson

### Technology

- 840 SHORT COMMUNICATION  
A new application of microwave technology to proteomics  
Hsueh-Fen Juan, Shing-Chuan Chang, Hsuan-Cheng Huang  
and Shui-Tein Chen
- 843 The development of an algorithm for the mass spectral interpretation  
of phosphoproteins  
Yupeng Zhao and Yen-Han Lin  
Supporting information see [www.proteomics-journal.de](http://www.proteomics-journal.de)
- 846 Tryptic transpeptidation products observed in proteome analysis  
by liquid chromatography-tandem mass spectrometry  
Heike Schaefer, Daniel C. Chamrad, Katrin Marcus, Kai A. Reidegeld,  
Martin Blüggel and Helmut E. Meyer



### Cover Illustration

has kindly been provided by  
Takao Kawakami,  
Clinical Proteome Center,  
Tokyo Medical University,  
Tokyo, Japan. This issue, p. 861

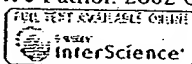
**Cellular location and age-dependent changes of the regulatory subunits of cAMP-dependent protein kinase in rat testis.**

Landmark BF, Oyen O, Skalhegg BS, Fauske B, Jahnsen T, Hausson V.

Institute of Medical Biochemistry, University of Oslo, Norway.

This study was undertaken to examine the expression and cellular location of the various cAMP-dependent protein kinase (PKA) subunits in different testicular cell types, using cDNA probes, isoenzyme-specific antibodies and activity measurements. Amounts of mRNA and protein were examined in cultured Sertoli cells, cultured peritubular cells, germ cells (pachytene spermatocytes, round spermatids), Leydig cell tumours as well as whole testes from rats of various ages. In Sertoli cells, there was a good correlation between the amount of mRNA and the respective immunoreactive proteins. In other types of cell, such as germ cells and Leydig tumour cells, this was not always the case. Large amounts of RII beta mRNA were found in Leydig tumour cells, whereas the amount of immunoreactive protein was low. Furthermore, large amounts of small-sized, germ cell-specific mRNAs for RI alpha (1.7 kb) and RII alpha (2.2 kb) were also found in the developing rat testis after 30 to 40 days of age, but the large amounts of mRNA were only partially reflected at the protein level. Pachytene spermatocytes and round spermatids were practically devoid of both RII alpha and RII beta protein. During spermatid differentiation, there was a decrease in RI alpha and an increase in RII alpha protein. Cell specific distribution of the various PKA subunits in testicular cell types is described. In some types of cell, discrepancies between mRNA and protein were demonstrated, which clearly suggest cell specific differences in translational efficiencies for some of these mRNAs, particularly the small-sized mRNAs for RI alpha and RII alpha in meiotic and post-meiotic germ cells.

PMID: 8107013 [PubMed - indexed for MEDLINE]



**Quantification of CK20 gene and protein expression in colorectal cancer by RT-PCR and immunohistochemistry reveals inter- and intratumour heterogeneity.**

Lassmann S, Bauer M, Soong R, Schreglmann J, Tabiti K, Nahrig J, Ruger R, Hofer H, Werner M.

Pathologisches Institut, Universitätsklinikum Freiburg, Albertstrasse 19, 79104 Freiburg, Germany. lassmann@ukl.uni-freiburg.de

Cytokeratin 20 (CK20) is an epithelial protein expressed almost exclusively in the gastrointestinal (GI) tract and is widely used as immunohistochemical marker for routine diagnosis. In contrast, CK20 gene expression is not an established marker for the classification of tumours and the detection of disseminated cancer cells in colorectal cancer. Recently, real-time reverse transcriptase polymerase chain reaction (RT-PCR) has provided the means for reproducible and quantitative investigation of molecular markers. This report directly compares CK20 mRNA and protein expression in serial sections of archival, formalin-fixed, paraffin-embedded (FFPE) colorectal adenocarcinomas. CK20 expression was detected by immunohistochemistry (IHC) in 60/63 (95.2%) cases, by conventional RT-PCR in 58/60 (96.7%) and by quantitative RT-PCR using the LightCycler (LightCycler is a trademark of a Member of the Roche Group) System in 29/32 (90.6%) microdissected cases, one case yielding variable results. Despite the high detection rate of all three techniques, marked heterogeneity of CK20 expression was seen between different cases and also within individual cases. CK20 expression profiles were not related to particular histopathological features of the tumours. A good correlation ( $r = 0.8964$ ) was found between CK20 mRNA and protein expression by comparing quantitative RT-PCR with IHC in 32 cases. This was also true for selected heterogeneous tumour cells within individual cases. Both RT-PCR and IHC are therefore valuable tools for CK20 detection in colorectal adenocarcinoma, with real-time RT-PCR providing supplementary quantitative information. This suggests a promising supportive role for quantitative RT-PCR in molecular pathology. Copyright 2002 John Wiley & Sons, Ltd.

Publication Types:

- Evaluation Studies

PMID: 12237879 [PubMed - indexed for MEDLINE]

**Expression of the multidrug resistance-associated protein (MRP) mRNA and protein in normal peripheral blood and bone marrow haemopoietic cells.**

**Legrand O, Perrot JY, Tang R, Simonin G, Gurbuxani S, Zittoun R, Marie JP.**

Laboratoire de Cinetique et de Cultures Cellulaires, Hotel Dieu, Paris, France.

We studied the expression of multidrug resistance-associated protein (MRP) in normal haemopoietic cells from peripheral blood and bone marrow. The MRP mRNA levels were estimated by RT/PCR and in situ hybridization (ISH) assay, and the protein levels by flow cytometry. 21 samples of peripheral blood and 21 samples of bone marrow (11 normal bone marrow donors, 10 patients in complete remission after chemotherapy for large cell lymphoma or acute myeloid leukaemia) were analysed. In peripheral blood the mean MRP mRNA level in CD3+ cells was statistically higher than in the other cells (3-fold by the methods used). The levels of MRP in CD3+ varied from one individual to another (4.5-34.8 units by RT/PCR and 5-23 grains/cell by ISH); however, this was proportional to the variation in all the cell lineages of same individual ( $r = 0.84$ ). In bone marrow the mean MRP levels of the various cell lineages (including CD34+) were similar to the basal level in HL60 cells. Individual expression levels were again variable; however, there was no difference between untreated normal bone marrow and post chemotherapy normal bone marrow. MRP protein expression was determined by flow cytometry with the monoclonal antibody MRPm6. The CD4+ lymphocytes exhibited a higher MRP protein expression than the other cell lineages, including CD8+ cells. There was a good correlation between the three methods used (RT/PCR and ISH,  $P = 0.0001$ ,  $r = 0.87$ ; RT/PCR and flow cytometry,  $P = 0.0001$ ,  $r = 0.85$ ; ISH and flow cytometry,  $P = 0.002$ ,  $r = 0.67$ ).

PMID: 8757504 [PubMed - indexed for MEDLINE]

149: Circulation. 2002 May 28;105(21):2524-30.

Related Articles, Links

Full text article at  
circ.ahajournals.org

**Vascular endothelial growth factor enhances cardiac allograft arteriosclerosis.**

Lemstrom KB, Krebs R, Nykanen AI, Tikkanen JM, Sihvola RK, Aaltola EM, Hayry PJ, Wood J, Alitalo K, Yla-Herttuala S, Koskinen PK.

Cardiopulmonary Research Group, Transplantation Laboratory, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland. Karl.Lemstrom@helsinki.fi

**BACKGROUND:** Cardiac allograft arteriosclerosis is a complex process of alloimmune response, chronic inflammation, and smooth muscle cell proliferation that includes cross talk between cytokines and growth factors. **METHODS AND RESULTS:** Our results in rat cardiac allografts established alloimmune response as an alternative stimulus capable of inducing vascular endothelial growth factor (VEGF) mRNA and protein expression in cardiomyocytes and graft-infiltrating mononuclear inflammatory cells, which suggests that these cells may function as a source of VEGF to the cells of coronary arteries. Linear regression analysis of these allografts with different stages of arteriosclerotic lesions revealed a strong correlation between intragraft VEGF protein expression and the development of intimal thickening, whereas blockade of signaling downstream of VEGF receptor significantly reduced arteriosclerotic lesions. In addition, in cholesterol-fed rabbits, intracoronary perfusion of cardiac allografts with a clinical-grade adenoviral vector that encoded mouse VEGF(164) enhanced the formation of arteriosclerotic lesions, possibly secondary to increased intragraft influx of macrophages and neovascularization in the intimal lesions. **CONCLUSIONS:** Our findings suggest a positive regulatory role between VEGF and coronary arteriosclerotic lesion formation in the allograft cytokine microenvironment.

PMID: 12034660 [PubMed - indexed for MEDLINE]

# GENIES

---

Benjamin Lewin

Oxford New York Tokyo  
Oxford University Press  
1997



## CHAPTER 29

# Regulation of transcription

The phenotypic differences that distinguish the various kinds of cells in a higher eukaryote are largely due to differences in the expression of genes that code for proteins, that is, those transcribed by RNA polymerase II. In principle, the expression of these genes might be regulated at any one of several stages. The concept of the "level of control" implies that gene expression is not necessarily an automatic process once it has begun. It could be regulated in a gene-specific way at any one of several sequential steps. We can distinguish (at least) five potential control points, forming the series:

Activation of gene structure  
↓  
Initiation of transcription  
↓  
Processing the transcript  
↓  
Transport to cytoplasm  
↓  
Translation of mRNA

The existence of the first step is implied by the discovery that genes may exist in either of two structural conditions. Relative to the state of most of the genome, genes are found in an "active" state in the cells in which they are expressed (see Chapter 27). The change of structure is distinct from the act of transcription, and indicates that the gene is "transcribable." This suggests that acquisition of the "active" structure must be the first step in gene expression.

Transcription of a gene in the active state is

controlled at the stage of initiation, that is, by the interaction of RNA polymerase with its promoter. This is now becoming susceptible to analysis in the *in vitro* systems (see Chapter 28). For most genes, this is a major control point; probably it is the most common level of regulation.

There is at present no evidence for control at subsequent stages of transcription in eukaryotic cells, for example, via antitermination mechanisms.

The primary transcript is modified by capping at the 5' end, and usually also by polyadenylation at the 3' end. Introns must be spliced out from the transcripts of interrupted genes. The mature RNA must be exported from the nucleus to the cytoplasm. Regulation of gene expression by selection of sequences at the level of nuclear RNA might involve any or all of these stages, but the one for which we have most evidence concerns changes in splicing; some genes are expressed by means of alternative splicing patterns whose regulation controls the type of protein product (see Chapter 30).

Finally, the translation of an mRNA in the cytoplasm can be specifically controlled. There is little evidence for the employment of this mechanism in adult somatic cells, but it does occur in some embryonic situations, as described in Chapter 7. The mechanism is presumed to involve the blocking of initiation of translation of some mRNAs by specific protein factors.

But having acknowledged that control of gene expression can occur at multiple stages, and that production of RNA cannot inevitably be equated with production of protein, it is clear

that the overwhelming majority of regulatory events occur at the initiation of transcription. Regulation of tissue-specific gene transcription lies at the heart of eukaryotic differentiation; indeed, we see examples in Chapter 38 in which proteins that regulate embryonic development prove to be transcription factors. A regulatory transcription factor serves to provide

common control of a large number of target genes, and we seek to answer two questions about this mode of regulation: what identifies the common target genes to the transcription factor; and how is the activity of the transcription factor itself regulated in response to intrinsic or extrinsic signals?

## Response elements identify genes under common regulation

The principle that emerges from characterizing groups of genes under common control is that they share a promoter element that is recognized by a regulatory transcription factor. An element that causes a gene to respond to such a factor is called a response element; examples are the HSE (heat shock response element), GRE (glucocorticoid response element), SRE (serum response element).

The properties of some inducible transcription factors and the elements that they recognize are summarized in Table 29.1. Response elements have the same general characteristics as upstream elements of promoters or enhancers. They contain short consensus sequences, and copies of the response elements found in different genes are closely related, but not necessarily identical. The region bound by the factor extends for a short distance on either side of

the consensus sequence. In promoters, the elements are not present at fixed distances from the startpoint, but are usually <200 bp upstream of it. The presence of a single element usually is sufficient to confer the regulatory response, but sometimes there are multiple copies.

Response elements may be located in promoters or in enhancers. Some types of elements are typically found in one rather than the other: usually an HSE is found in a promoter, while a GRE is found in an enhancer. We assume that all response elements function by the same general principle. A gene is regulated by a sequence at the promoter or enhancer that is recognized by a specific protein. The protein functions as a transcription factor needed for RNA polymerase to initiate. Active protein is available only under conditions when the gene is to be expressed; its absence means that the promoter is not activated by this particular circuit.

An example of a situation in which many genes are controlled by a single factor is provided by the heat shock response. This is common to a wide range of prokaryotes and eukaryotes and involves multiple controls of gene expression: an increase in temperature turns off transcription of some genes, turns on transcription of the heat shock genes, and causes changes in the translation of mRNAs. The control of the heat shock genes illustrates the differences between prokaryotic and eukaryotic modes of control. In bacteria, a new sigma factor is synthesized that directs RNA polymerase holoenzyme to recognize an alter-

**Table 29.1** Inducible transcription factors bind to response elements that identify groups of promoters or enhancers subject to coordinate control.

Regulatory Agent	Module	Consensus	Factor
Heat shock	HSE	CNNGAANNTCGNG	HSTF
Glucocorticoid	GRE	TGGTACAAATGTTCT	Receptor
Phorbol ester	TRE	TGACTCA	AP1
Serum	SRE	CCATATTAGG	SRE



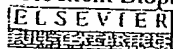
## Retinal preconditioning and the induction of heat-shock protein 27.

Li Y, Roth S, Laser M, Ma JX, Crosson CE.

Department of Ophthalmology, Medical University of South Carolina, Charleston, South Carolina 29425, USA.

**PURPOSE:** Brief periods of ischemia have been shown to protect the retina from potentially damaging periods of ischemia. This phenomenon has been termed ischemic preconditioning or ischemic tolerance. In the present study the cellular changes in levels of heat shock protein (Hsp)27, -70, and -90 mRNA and expression of Hsp in the rat retina associated with ischemic preconditioning were evaluated. **METHODS:** Unilateral retinal ischemia was created in Long-Evans and Sprague-Dawley rats for 5 minutes. Rats were then left for 1 hour to 7 days, to allow the retina to reperfuse. Retinas were dissected, the mRNA and protein isolated, and Northern and Western blot analyses conducted to detect changes in expression of Hsp27, -70, and -90. Immunohistochemical studies were used to identify retinal regions where Hsp changes occurred. Selected animals were subjected to a second ischemic event, 60 minutes in duration, to correlate the changes in expression of Hsp with functional protection of the retina from ischemic injury. **RESULTS:** In control and sham-treated animals retinal Hsp27, -70, and -90 mRNAs were detectable. Five hours after retinal preconditioning, levels of Hsp27 mRNA were elevated above control levels, and 24 hours later, mRNA levels increased 200% over basal levels. Hsp27 expression remained elevated for up to 72 hours and then began to return to control levels. Hsp27 protein levels were increased by 200% over basal levels 24 hours after retinal preconditioning, remained at this level for 72 hours, and then returned to control levels. In contrast, no consistent change in Hsp70 or -90 mRNA or protein levels was observed during the course of the study. Immunohistochemical studies demonstrated that the increase in expression of Hsp27 was localized to neuronal and non-neuronal cells in the inner layers of the retina. Electoretinography studies demonstrated a strong correlation between the protection of retinal function from ischemic injury and the expression of Hsp27. **CONCLUSIONS:** These results provide evidence that the induction of Hsp27 is a gene-specific event associated with ischemic preconditioning in the retina. This increase in expression of Hsp27 occurs in both neuronal and non-neuronal retinal cells, and appears to be one component of the neuroprotective events induced by ischemic preconditioning in the retina.

PMID: 12601062 [PubMed - indexed for MEDLINE]



**Enhanced expressions of arachidonic acid-sensitive tandem-pore domain potassium channels in rat experimental acute cerebral ischemia.**

**Li ZB, Zhang HX, Li LL, Wang XL.**

Institute of Materia Medica, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100050, China.

To further explore the pathophysiological significance of arachidonic acid-sensitive potassium channels, RT-PCR and Western blot analysis were used to investigate the expression changes of TREK channels in cortex and hippocampus in rat experimental acute cerebral ischemia in this study. Results showed that TREK-1 and TRAAK mRNA in cortex, TREK-1 and TREK-2 mRNA in hippocampus showed significant increases 2 h after middle cerebral artery occlusion (MCAO). While the mRNA expression levels of the all three channel subtypes increased significantly 24 h after MCAO in cortex and hippocampus. At the same time, the protein expressions of all the three channel proteins showed significant increase 24 h after MCAO in cortex and hippocampus, but only TREK-1 showed increased expression 2 h after MCAO in cortex and hippocampus. Immunohistochemical experiments verified that all the three channel proteins had higher expression levels in cortical and hippocampal neurons 24 h after MCAO. These results suggested a strong correlation between TREK channels and acute cerebral ischemia. TREK channels might provide a neuroprotective mechanism in the pathological process.

PMID: 15652517 [PubMed - indexed for MEDLINE]

ELSEVIER  
SCIENCE

**Increasing expression of tissue plasminogen activator and plasminogen activator inhibitor type 2 in dog gingival tissues with progressive inflammation.**

Lindberg P, Kinnby B, Lecander I, Lang NP, Matsson L.

Center for Oral Health Sciences, Malmo University, S-214 21 Malmo, Sweden.  
pia.lindberg@od.mah.se

Urokinase and tissue-type plasminogen activators (u-PA and t-PA) are serine proteases that convert plasminogen into plasmin, which degrades matrix proteins and activates metalloproteinases. The PAs are balanced by specific inhibitors (PAI-1 and PAI-2). Local production of t-PA and PAI-2 was recently demonstrated in human gingival tissues. The aim now was to investigate the production and localization of t-PA and PAI-2 in gingival tissues from dogs in three well-defined periodontal conditions; clinically healthy gingiva, chronic gingivitis and an initial stage of ligature-induced loss of attachment. At the start of the experiment the gingiva showed clear signs of inflammation. Clinically healthy gingiva were obtained after 21 days period of intense oral hygiene. Attachment loss was induced by placing rubber ligatures around the neck of some teeth. Biopsies were taken from areas representing the different conditions and prepared for in situ hybridization and immunohistochemistry. In clinically healthy gingiva both t-PA mRNA and antigen were expressed in a thin outer layer of the sulcular and junctional epithelia. No t-PA signals or staining were seen in connective tissue. Both mRNA signaling and immunostaining for t-PA were stronger in chronic gingivitis. In areas with loss of attachment, t-PA mRNA as well as antigen were found in the sulcular and junctional epithelia to a similar degree as in gingivitis. Occasionally the connective tissue was involved, especially in connection with vessels. PAI-2 mRNA was seen in a thin outer layer of the sulcular and junctional epithelia in clinically healthy gingiva, but no signals were seen in connective tissue. PAI-2 antigen was found primarily in the outer layer of the sulcular and junctional epithelia. Some cells in the connective tissue were stained. In gingivitis, PAI-2 signals were mainly found in the same locations, but more intense and extending towards the connective tissue. Immunostaining was seen in the outer half of the sulcular and junctional epithelia as well as in the upper part of the connective tissue, close to the sulcular epithelium. In sites with loss of attachment, PAI-2 mRNA was found throughout the sulcular and junctional epithelia, as was the antigen, which stained intensely. No PAI-2 mRNA was seen in connective tissue; the antigen was found scattered, especially near vessels. This study shows that the expression of both t-PA and PAI-2 increases with experimental gingival inflammation in the dog, and furthermore, the two techniques demonstrate a strong correlation between the topographical distribution of the site of protein synthesis and the tissue location of the antigens for both t-PA and PAI-2. The distribution correlates well with previous findings in humans.

**Effect of duration of fixation on quantitative reverse transcription polymerase chain reaction analyses.**

Macabeo-Ong M, Ginzinger DG, Dekker N, McMillan A, Regezi JA, Wong DT, Jordan RC.

Oral Pathology, Department of Stomatology, University of California San Francisco, California 94143-0424, USA.

Increasingly, there is the need to analyze gene expression in tumor tissues and correlate these findings with clinical outcome. Because there are few tissue banks containing enough frozen material suitable for large-scale genetic analyses, methods to isolate and quantify messenger RNA (mRNA) from formalin-fixed, paraffin-embedded tissue sections are needed. Recovery of RNA from routinely processed biopsies and quantification by the polymerase chain reaction (PCR) has been reported; however, the effects of formalin fixation have not been well studied. We used a proteinase K-salt precipitation RNA isolation protocol followed by TaqMan quantitative PCR to compare the effect of formalin fixation for 24, 48, and 72 hours and for 1 week in normal (2), oral epithelial dysplasia (3), and oral squamous cell carcinoma (4) specimens yielding 9 fresh and 36 formalin-fixed samples. We also compared mRNA and protein expression levels using immunohistochemistry for epidermal growth factor receptor (EGFR), matrix metalloproteinase (MMP)-1, p21, and vascular endothelial growth factor (VEGF) in 15 randomly selected and routinely processed oral carcinomas. We were able to extract RNA suitable for quantitative reverse transcription (RT) from all fresh (9/9) and formalin-fixed (36/36) specimens fixed for differing lengths of time and from all (15/15) randomly selected oral squamous cell carcinoma. We found that prolonged formalin fixation (>48 h) had a detrimental effect on quantitative RT polymerase chain reaction results that was most marked for MMP-1 and VEGF but less evident for p21 and EGFR. Comparisons of quantitative RT polymerase chain reaction and immunohistochemistry showed that for all markers, except p21, there was good correlation between mRNA and protein levels. p21 mRNA was overexpressed in only one case, but protein levels were elevated in all but one tumor, consistent with the established translational regulation of p21. These results show that RNA can be reliably isolated from formalin-fixed, paraffin-embedded tissue sections and can produce reliable quantitative RT-PCR data. However, results for some markers are adversely affected by prolonged formalin fixation times.

PMID: 12218216 [PubMed - indexed for MEDLINE]



**Id-1 and Id-2 are overexpressed in pancreatic cancer and in dysplastic lesions in chronic pancreatitis.**

Maruyama H, Kleeff J, Wildt S, Friess H, Buchler MW, Israel MA, Kore M.

Division of Endocrinology, Department of Medicine, University of California, Irvine, USA.

Id proteins antagonize basic helix-loop-helix proteins, inhibit differentiation, and enhance cell proliferation. In this study we compared the expression of Id-1, Id-2, and Id-3 in the normal pancreas, in pancreatic cancer, and in chronic pancreatitis (CP). Northern blot analysis demonstrated that all three Id mRNA species were expressed at high levels in pancreatic cancer samples by comparison with normal or CP samples. Pancreatic cancer cell lines frequently coexpressed all three Ids, exhibiting a good correlation between Id mRNA and protein levels, as determined by immunoblotting with highly specific anti-Id antibodies. Immunohistochemistry using these antibodies demonstrated the presence of faint Id-1 and Id-2 immunostaining in pancreatic ductal cells in the normal pancreas, whereas Id-3 immunoreactivity ranged from weak to strong. In the cancer tissues, many of the cancer cells exhibited abundant Id-1, Id-2, and Id-3 immunoreactivity. Scoring on the basis of percentage of positive cells and intensity of immunostaining indicated that Id-1 and Id-2 were increased significantly in the cancer cells by comparison with the respective controls. Mild to moderate Id immunoreactivity was also seen in the ductal cells in the CP-like areas adjacent to these cells and in the ductal cells of small and interlobular ducts in CP. In contrast, in dysplastic and atypical papillary ducts in CP, Id-1 and Id-2 immunoreactivity was as significantly elevated as in the cancer cells. These findings suggest that increased Id expression may be associated with enhanced proliferative potential of pancreatic cancer cells and of proliferating or dysplastic ductal cells in CP.

PMID: 10487839 [PubMed - indexed for MEDLINE]

## Id-1 and Id-2 Are Overexpressed in Pancreatic Cancer and in Dysplastic Lesions in Chronic Pancreatitis

Haruhisa Maruyama,\* Jörg Kleeff,\* Stefan Wildi,\*  
Helmut Friess,<sup>†</sup> Markus W. Büchler,<sup>†</sup>  
Mark A. Israel,<sup>‡</sup> and Murray Korc\*

From the Division of Endocrinology, Diabetes, and Metabolism,\*  
Departments of Medicine, Biological Chemistry and  
Pharmacology, University of California, Irvine, California; the  
Department of Visceral and Transplantation Surgery,<sup>†</sup> University  
of Bern, Bern, Switzerland; and the Preuss Laboratory,<sup>‡</sup>  
Department of Neurological Surgery, University of California,  
San Francisco, California

Id proteins antagonize basic helix-loop-helix proteins, inhibit differentiation, and enhance cell proliferation. In this study we compared the expression of Id-1, Id-2, and Id-3 in the normal pancreas, in pancreatic cancer, and in chronic pancreatitis (CP). Northern blot analysis demonstrated that all three Id mRNA species were expressed at high levels in pancreatic cancer samples by comparison with normal or CP samples. Pancreatic cancer cell lines frequently coexpressed all three Ids, exhibiting a good correlation between Id mRNA and protein levels, as determined by immunoblotting with highly specific anti-Id antibodies. Immunohistochemistry using these antibodies demonstrated the presence of faint Id-1 and Id-2 immunostaining in pancreatic ductal cells in the normal pancreas, whereas Id-3 immunoreactivity ranged from weak to strong. In the cancer tissues, many of the cancer cells exhibited abundant Id-1, Id-2, and Id-3 immunoreactivity. Scoring on the basis of percentage of positive cells and intensity of immunostaining indicated that Id-1 and Id-2 were increased significantly in the cancer cells by comparison with the respective controls. Mild to moderate Id immunoreactivity was also seen in the ductal cells in the CP-like areas adjacent to these cells and in the ductal cells of small and interlobular ducts in CP. In contrast, in dysplastic and atypical papillary ducts in CP, Id-1 and Id-2 immunoreactivity was as significantly elevated as in the cancer cells. These findings suggest that increased Id expression may be associated with enhanced proliferative potential of pancreatic cancer cells and of proliferating or dysplastic ductal cells in CP. (*Am J Pathol* 1999, 155:815-822)

Basic helix-loop-helix (bHLH) proteins play an important role as transcription factors in cellular development, proliferation, and differentiation.<sup>1,2</sup> The basic domain of the bHLHs is required for binding to an E-box DNA sequence, thus promoting transcription of specific target genes. The HLH domain promotes dimer formation with various members of the bHLH protein family.<sup>1,2</sup> Homodimers of the class B family of bHLH proteins, including MyoD, NeuroD, and numerous other proteins, are known to activate tissue-specific genes.<sup>3-5</sup> These tissue-specific bHLHs typically form heterodimers with widely expressed class A bHLHs, which include proteins encoded by E2A, E2-2, HEB, and other genes (also termed E-proteins).<sup>6-9</sup> These heterodimers activate transcription of genes that are associated with differentiation.

Id genes encode a family of four HLH proteins that lack the basic DNA binding domain.<sup>1,10</sup> They act as dominant-negative HLH proteins by forming high affinity heterodimers with other bHLH proteins, thereby preventing them from binding to DNA and inhibiting transcription of differentiation-associated genes.<sup>10-12</sup> Id gene expression is down-regulated on differentiation in many cell types *in vitro* and *in vivo*.<sup>13-18</sup> In addition, Id proteins seem to be required for cell cycle progression through G<sub>1</sub>/S phase in certain cell types, and interaction between Id-2 and pRB is associated with enhanced proliferation in some cell lines *in vitro*.<sup>19-23</sup>

Pancreatic cancer is the fifth leading cause of cancer death in the United States, with a mortality rate that virtually equals its incidence rate.<sup>24</sup> This malignancy is often associated with the overexpression of a variety of mitogenic growth factors and their receptors, and by oncogenic mutations of K-ras and inactivation of the p53 tumor suppressor gene.<sup>25</sup> We have recently reported that pancreatic cancers overexpress the HLH protein Id-2, and that enhanced expression of this protein is evident in the cytoplasm of the cancer cells within the pancreatic tumor mass.<sup>26</sup> It is not known, however, whether the expression of other Id proteins is altered in this malignancy, or whether their expression is altered in chronic pancreatitis

Contract grant sponsor: National Cancer Institute. Contract grant number: U. S. Public Health Service grant CA-40162.

Accepted for publication May 24, 1999.

Address reprint requests to Dr. Murray Korc, Division of Endocrinology, Diabetes and Metabolism, Medical Sciences I, C240, University of California, Irvine, CA 92697. E-mail: mkorc@uci.edu.



(CP), an inflammatory disease that is characterized by dysplastic ducts, foci of proliferating ductal cells, acinar cell degeneration, and fibrosis.<sup>27</sup> We now report that there is a five- to sixfold increase in Id-1 and Id-2 mRNA levels and a twofold increase in Id-3 mRNA levels in pancreatic cancer by comparison with the normal pancreas. In contrast, overall Id mRNA levels are not increased in CP.

### Patients and Methods

Normal human pancreatic tissue samples from 7 male and 5 female donors (median age 41.8 years, range 14–68 years), CP tissues from 13 males and 1 female (median age 42.1 years; range 30–56 years), and pancreatic cancer tissues from 10 male and 6 female donors (median age 62.6 years; range 53–83 years) were obtained through an organ donor program and from surgical specimens from patients with severe symptomatic chronic pancreatitis or pancreatic cancer. A partial duodenopancreatectomy (Whipple/pylorus-preserving Whipple;  $n = 13$ ), a left resection of the pancreas ( $n = 2$ ), or a total pancreatectomy ( $n = 1$ ) were carried out in the pancreatic cancer patients. According to the TNM classification of the Union Internationale Contre le Cancer (UICC) 6 tumors were stage 1, 1 was stage 2, and 9 were stage 3 ductal cell adenocarcinoma. Freshly removed tissue samples were fixed in 10% formaldehyde solution for 12 to 24 hours and paraffin-embedded for histological analysis. In addition, tissue samples were frozen in liquid nitrogen immediately on surgical removal and maintained in  $-80^{\circ}\text{C}$  until use for RNA extraction. All studies were approved by the Ethics Committee of the University of Bern, Bern, Switzerland, and by the Human Subjects Committee at the University of California, Irvine, California.

### Northern Blot Analysis

Northern blot analysis was carried out as described previously.<sup>26,28</sup> Briefly, total RNA was extracted by the single step acid guanidinium thiocyanate phenol chloroform method. RNA was size-fractionated on 1.2% agarose/1.8 mol/L formaldehyde gels; electrotransferred onto nylon membranes, and cross-linked by UV irradiation. Blots were prehybridized and hybridized with cDNA probes and washed under high stringency conditions. The following cDNA probes were used: a 979-bp human Id-1 cDNA probe, a 440-bp human Id-2 cDNA probe, and a 450-bp human Id-3 cDNA probe, covering the entire coding regions of Id-1, Id-2, and Id-3, respectively. A *Bam*HI 190-bp fragment of mouse 7S cDNA that hybridizes with human cytoplasmic RNA was used to confirm equal RNA loading and transfer. Blots were then exposed at  $-80^{\circ}\text{C}$  to Kodak BioMax-MS films and the resulting autoradiographs were scanned to quantify the intensity of the radiographic bands.<sup>26,28</sup> For each sample the ratio of Id mRNA expression to 7S expression was calculated. To compare the relative increase in expression of the respective Id mRNA species in the cancer and CP samples, the same normal samples were used for normal/

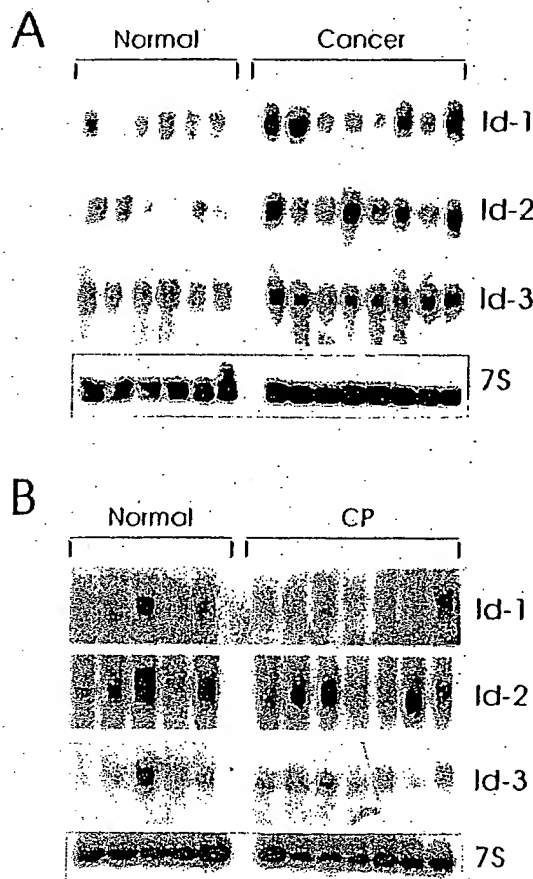


Figure 1. mRNA expression of Id-1, Id-2, and Id-3 in pancreatic cancer and chronic pancreatitis. Total RNA (20  $\mu\text{g}/\text{lane}$ ) from six normal, eight cancerous, and seven chronic pancreatitis tissue samples were subjected to Northern blot analysis using  $^{32}\text{P}$ -labeled cDNA probes (500,000 cpm/ml) specific for Id-1, Id-2, and Id-3, respectively. A 7S cDNA probe (50,000 cpm/ml) was used as a loading and transfer control. Exposure times of the normal/cancer blots were 1 day for all Id probes, and 2 days for the normal/CP blots. Exposure time was 4 hours for mouse 7S cDNA. By comparison with the normal samples, Id-1 and Id-3 mRNA levels were elevated in 8 and 9 cancer samples, respectively, whereas Id-2 was elevated in 6 cancer samples.

cancer and normal/CP membranes. The median score for Id-1, Id-2, and Id-3 mRNA levels in these normal samples was set to 100. Statistical analysis was performed with SigmaStat software (Jandel Scientific, San Raphael, CA). The rank sum test was used, and  $P < 0.05$  was taken as the level of significance.

### Cell Culture and Western Blot Analysis

PANC-1, MIA-PaCa-2, ASPC-1, and CAPAN-1 human pancreatic cell lines were obtained from ATCC (Manassas, VA). COLO-357 human pancreatic cells were a gift from Dr. R. S. Metzger (Durham, NC). Cells were routinely grown in DMEM (COLO-357, MIA-PaCa-2, PANC-1) or RPMI (ASPC-1, CAPAN-1) supplemented with 10% fetal bovine serum, 100 U/ml penicillin, and 100  $\mu\text{g}/\text{ml}$  streptomycin. For immunoblot analysis, exponentially growing

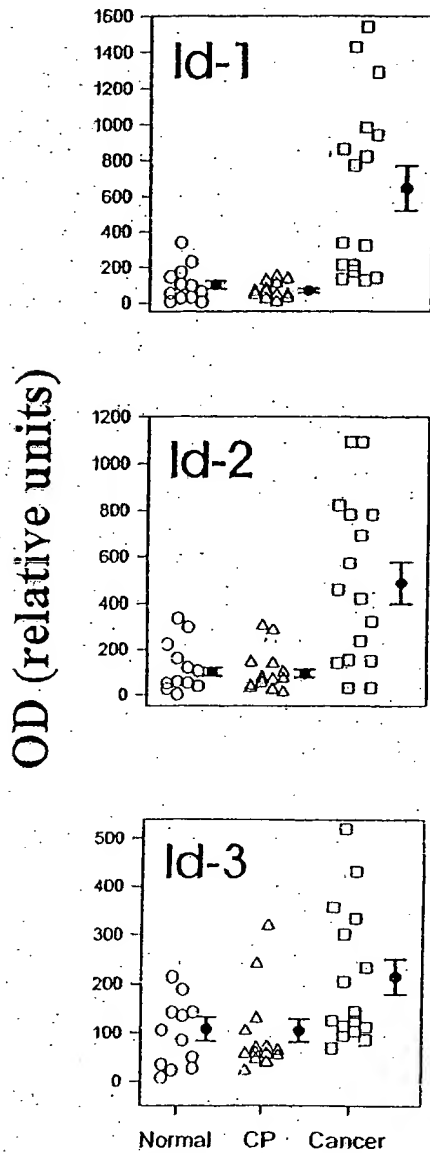


Figure 2. Densitometric analysis of Northern blots. Autoradiographs of Northern blots from 12 normal, 14 CP, and 16 pancreatic cancers were analyzed by densitometry. mRNA levels were determined by calculating the ratio of the optical density for the respective Id mRNA species in relation to the optical density of mouse 7S cDNA. To compare the relative increase in expression of the respective Id mRNA species in the cancer and CP samples, the same normal samples were used for normal/cancer and normal/CP membranes. Normal pancreatic tissues are indicated by circles, CP tissues by triangles, and cancer tissues by squares. Data are expressed as median scores  $\pm$  SD. By comparison with the normal samples, only the cancer samples exhibited significant increases: 6.5-fold ( $P < 0.01$ ) for Id-1, fivefold ( $P < 0.01$ ) for Id-2, and twofold ( $P = 0.027$ ) for Id-3.

cells (60–70% confluent) were solubilized in lysis buffer containing 50 mmol/L Tris-HCl, pH 7.4, 150 mmol/L NaCl, 1 mmol/L EDTA, 1  $\mu$ g/ml pepstatin A, 1 mmol/L phenylmethylsulfonyl fluoride (PMSF), and 1% Triton X-100. Proteins were subjected to sodium dodecyl sulfate polyacryl-

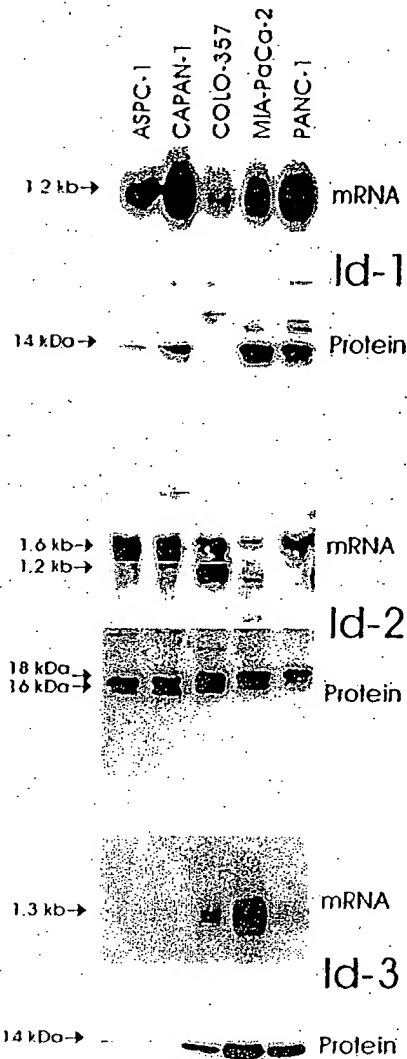
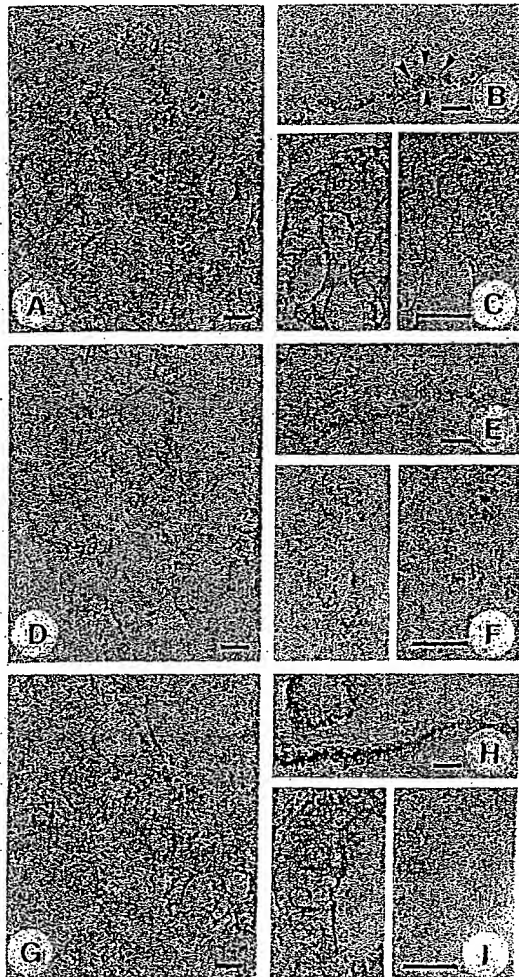


Figure 3. Id mRNA and protein expression in pancreatic cancer cell lines. Upper panels: Total RNA (20  $\mu$ g/lane) from 5 pancreatic cancer cell lines were subjected to Northern blot analysis using  $^{32}$ P-labeled cDNA probes (500,000 cpm/ml) specific for Id-1, Id-2, and Id-3, respectively. Exposure times were 1 day for all Id probes. Lower panels: Immunoblotting. Cell lysates (30  $\mu$ g/lane) were subjected to SDS-PAGE. Membranes were probed with specific Id-1, Id-2, and Id-3 antibodies. Visualization was performed by enhanced chemiluminescence.

amide gel electrophoresis (SDS-PAGE), transferred to Immobilon P membranes, and incubated for 90 minutes with the indicated antibodies and for 60 minutes with secondary antibodies against rabbit IgG. Visualization was performed by enhanced chemiluminescence.

#### Immunohistochemistry

Specific rabbit anti-human Id-1 (C-20), Id-2 (C-20), and Id-3 (C-20; all from Santa Cruz Biotechnology, Santa



**Figure 4.** Normal and cancerous pancreatic tissues were subjected to immunostaining using highly specific anti-Id-1 (A-C), anti-Id-2 (D-F), and anti-Id-3 (G-I) antibodies as described in the Methods section. Moderate to strong Id-1 immunoreactivity was present in the cytoplasm of duct-like cancer cells (A and C, left panel). In the normal pancreas there was weak Id-1 immunoreactivity in the ductal cells (B). Preabsorption with the Id-1-specific blocking peptide abolished the Id-1 immunoreactivity (C, right panel). Strong Id-2 immunoreactivity was observed in the cytoplasm of the cancer cells that exhibited duct-like structures (D and F, left panel), whereas in the normal pancreas, there was only weak Id-2 immunoreactivity in the ductal cells (E). Preabsorption with the Id-2-specific blocking peptide abolished the Id-2 immunoreactivity (F, right panel). Moderate to strong Id-3 immunoreactivity was present in the duct-like cancer cells (G and I, left panel). Moderate to strong Id-3 immunoreactivity was also present in the ductal cells of normal pancreatic tissue samples (H). Id-3 immunoreactivity was completely abolished by preabsorption with the Id-3 specific blocking peptide (I, right panel). A, D, and G constitute serial sections of a pancreatic cancer sample, revealing coexpression of the three Id proteins. Scale bars, 25  $\mu$ m.

Cruz, CA) polyclonal antibodies were used for immunohistochemistry. These affinity-purified rabbit polyclonal antibodies specifically react with Id-1, Id-2, and Id-3, respectively, of human origin, as determined by Western blotting. Paraffin-embedded sections (4  $\mu$ m) were subjected to immunostaining using the streptavidin-peroxidase technique. Where indicated, immunostaining for all three Id proteins was performed on serial sections. En-

dogenous peroxidase activity was blocked by incubation for 30 minutes with 0.3% hydrogen peroxide in methanol. Tissue sections were incubated for 15 minutes (23°C) with 10% normal goat serum and then incubated for 16 hours at 4°C with the indicated antibodies in PBS containing 1% bovine serum albumin. Bound antibodies were detected with biotinylated goat anti-rabbit IgG secondary antibodies and streptavidin-peroxidase complex, using diaminobenzidine tetrahydrochloride as the substrate. Sections were counterstained with Mayer's hematoxylin. Preabsorption with Id-1-, Id-2-, or Id-3-specific blocking peptides completely abolished immunoreactivity of the respective primary antibody. The immunohistochemical results were semiquantitatively analyzed as described previously.<sup>29,30</sup> The percentage of positive cancer cells was stratified into four groups: 0, no cancer cells exhibiting immunoreactivity; 1, <33% of the cancer cells exhibiting immunoreactivity; 2, 33 to 67% of the cancer cells exhibiting immunoreactivity; 3 >67% of the cancer cells exhibiting immunoreactivity. The intensity of the immunohistochemical signal was also stratified into four groups: 0, no immunoreactivity; 1, weak immunoreactivity; 2, moderate immunoreactivity; 3, strong immunoreactivity. Finally, the sum of the results of the cell score and the intensity score was calculated. Statistical analysis was performed with SigmaStat software. The rank sum test was used, and  $P < 0.05$  was taken as the level of significance.

## Results

Northern blot analysis of total RNA isolated from 12 normal pancreatic tissues and 16 pancreatic cancers revealed the presence of the 1.2-kb Id-1 transcript and the 1.6-kb Id-2 mRNA transcript in 11 of the 12 normal pancreatic samples, and the 1.3-kb Id-3 mRNA transcript in all normal pancreatic samples (Figure 1A, 2). In the cancer tissues, Id-1 mRNA levels were elevated in 8 of 16 samples, Id-2 mRNA levels were elevated in 9 of these samples, and Id-3 mRNA levels were elevated in 6 of these samples (Figure 1A, 2). Concomitant overexpression of all three Id species was observed in 6 of the cancer samples (38%). In contrast, none of the Id mRNA species were overexpressed in CP by comparison with normal controls (Figure 1B, 2). Densitometric analysis of all of the autoradiograms indicated that there was a 6.5-fold increase ( $P < 0.01$ ) in Id-1 mRNA levels, a fivefold increase ( $P < 0.01$ ) in Id-2 mRNA levels, and a twofold increase ( $P = 0.027$ ) in Id-3 mRNA levels in the pancreatic cancer samples in comparison to normal controls (Figure 2). In contrast, there was no statistically significant difference in the expression levels of Id-1, Id-2, and Id-3, in CP tissues in comparison to the corresponding levels in the normal pancreas (Figure 2).

Next, we assessed the expression of the three Id genes in 5 human pancreatic cancer cell lines by Northern and Western blot analyses. Id-1 mRNA was present at varying levels in all 5 cell lines (Figure 3). ASPC-1, CAPAN-1, MIA-PaCa-2, and PANC-1 expressed moderate to high levels of Id-1 mRNA, whereas COLO-357 cells

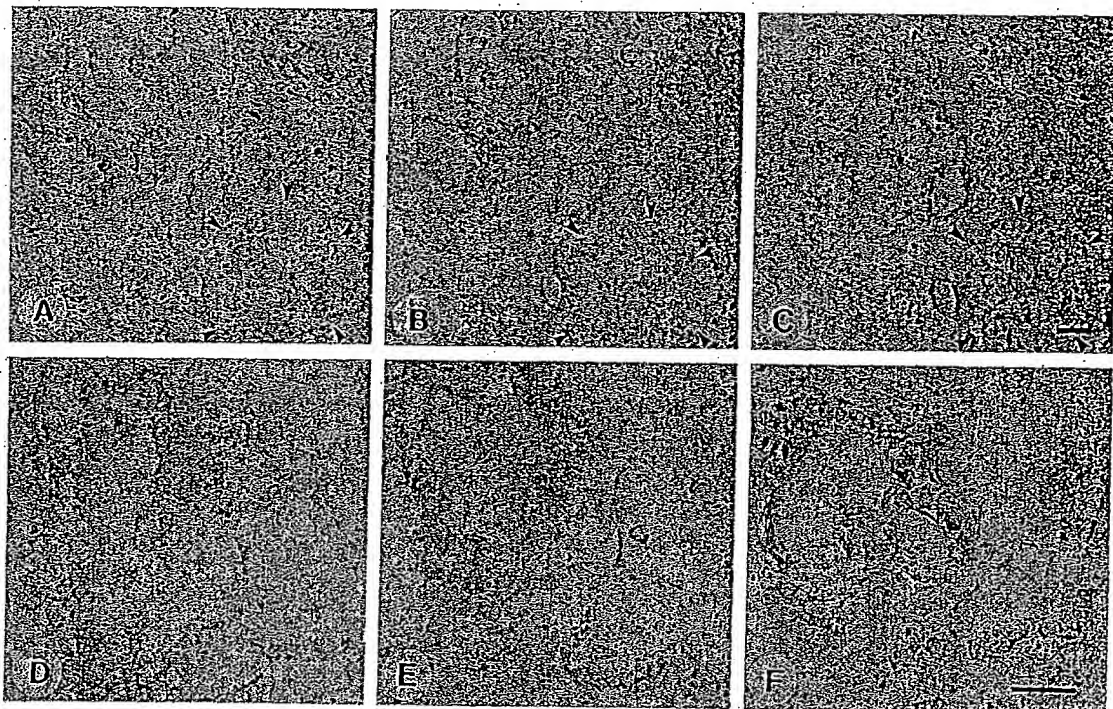


Figure 5. Immunohistochemistry of pancreatic cancer and dysplastic ducts in CP tissues. In the pancreatic cancer tissues (A-C) there was moderate to strong Id-1 (A), Id-2 (B), and Id-3 (C) immunoreactivity in the ductal cells in the areas adjacent to the cancer cells that exhibited CP-like alterations. Islet cells did not exhibit Id immunoreactivity (outlined by solid arrowheads). In the CP samples, moderate to strong Id-1 (D), Id-2 (E), and Id-3 (F) immunoreactivity was present in the cytoplasm of epithelial cells forming large dysplastic ducts. Scale bar, 25  $\mu$ m.

expressed relatively low levels of this mRNA moiety. Western blotting with a highly specific anti-Id-1 antibody confirmed the presence of the approximately 14-kd Id-1 protein in the 4 cell lines that expressed high levels of Id-1 mRNA (Figure 3). Furthermore, the three cell lines with the highest Id-1 mRNA expression (CAPAN-1, MIA-PaCa-2, and PANC-1) also exhibited the highest Id-1 protein expression. Variable levels of the 1.6-kb Id-2 mRNA transcript were present in all 5 cell lines. In addition, a minor band of approximately 1.2 kb was visible in COLO-357 and MIA-PaCa-2 cells. Immunoblot analysis with a highly specific anti-Id-2 antibody revealed two bands of approximately 16 and 18 kd at relatively high levels in all of the cell lines with exception of PANC-1 cells, in which the 16-kd band was relatively faint (Figure 3). With the exception of MIA-PaCa-2 cells, there was a good correlation between Id-2 mRNA and protein levels (Figure 3). Id-3 mRNA was present at high levels in MIA-PaCa-2 cells, at moderate levels in COLO-357 cells, and at low levels in PANC-1 cells. Id-3 mRNA was not detectable in ASPC-1 and CAPAN-1 cells (Figure 3). Immunoblot analysis with a highly specific anti-Id-3 antibody revealed an approximately 14-kd band that was most abundant in MIA-PaCa-2 cells, and was also readily apparent in COLO-357 and PANC-1 cells. In contrast, only a faint Id-3 band was seen in ASPC-1 and CAPAN-1 cells. Thus, with the exception of PANC-1 cells, there was a good correlation between Id-3 mRNA and protein levels.

To determine the localization of Id-1, Id-2, and Id-3, immunostaining was carried out using the same highly specific anti-Id antibodies. In the pancreatic cancers, moderate to strong Id-1 immunoreactivity was present in the cancer cells in 9 of 10 randomly selected cancer samples. An example of moderate Id-1 immunoreactivity is shown in Figure 4A, and of strong immunoreactivity in Figure 4C (left panel). In contrast, in the normal pancreas, faint Id-1 immunoreactivity was present only in the ductal cells of pancreatic ducts (Figure 4B, arrowheads). Preabsorption with the Id-1-specific blocking peptide completely abolished the Id-1 immunoreactivity (Figure 4C, right panel). The cancer cells also exhibited strong Id-2 (Figure 4, D and F, left panel) and moderate to strong Id-3 immunoreactivity. An example of moderate Id-3 immunoreactivity is shown in Figure 4G, and of strong immunoreactivity in Figure 4I (left panel). In contrast, only faint Id-2 immunoreactivity was present in the ductal cells in the normal pancreas (Figure 4E), whereas Id-3 immunoreactivity in these cells was more variable and ranged from moderate to occasionally strong (Figure 4H). Islet cells and acinar cells were always devoid of Id immunoreactivity. Preabsorption of the respective antibody with the blocking peptides specific for Id-2 (Figure 4F, right panel) and Id-3 (Figure 4I, right panel) completely abolished immunoreactivity. Analysis of serial pancreatic cancer sections revealed that there was often colocalization of the

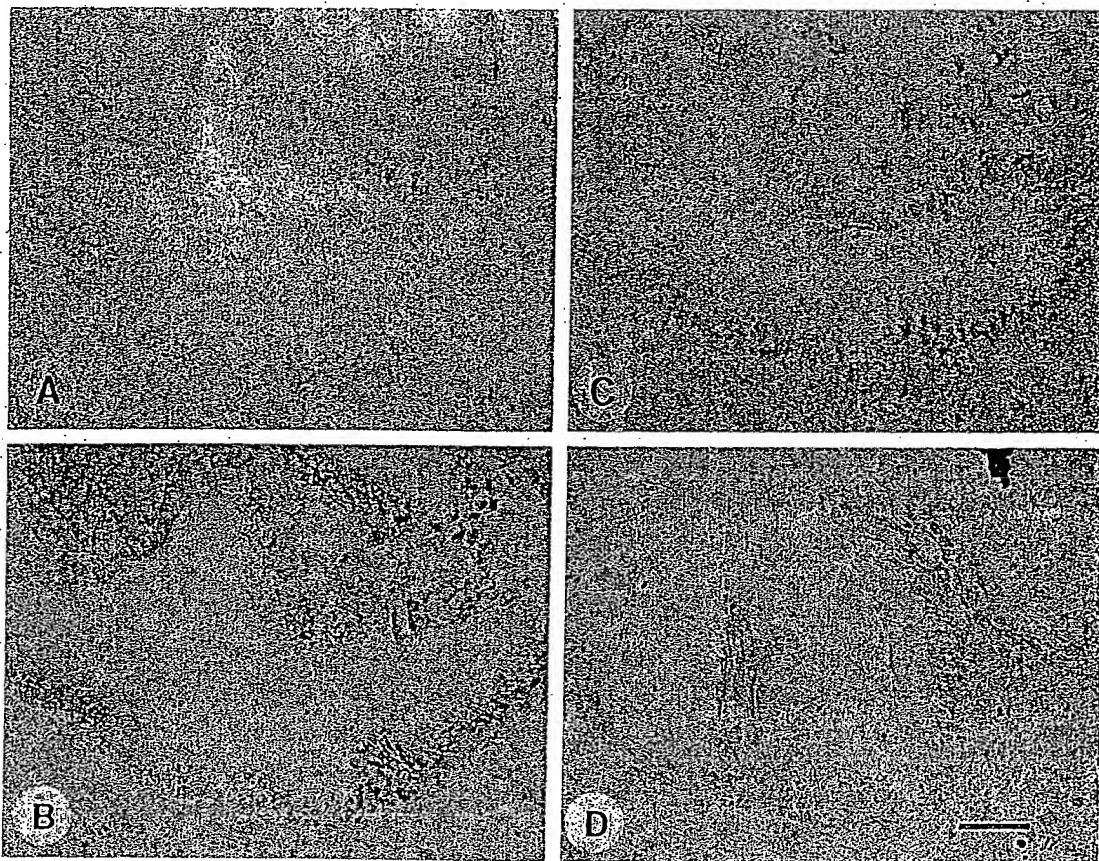


Figure 6. Immunohistochemistry of atypical papillary epithelium in CP tissues. Serial section analysis of some CP samples revealed the presence of large duct-like structures with atypical papillary epithelium. Mild to moderate Id-1 (A) and Id-2 (B) immunoreactivity and weak Id-3 (C) immunoreactivity was present in the cytoplasm of the cells forming these large ducts with papillary structures. Some CP samples also exhibited moderate Id-3 immunoreactivity in these cells (D). Scale bar, 25  $\mu$ m.

three Id proteins. An example of serial sections from a pancreatic cancer tissue is shown in Figure 4, A, D, and G.

Id-1, Id-2, and Id-3 immunoreactivity was also present at moderate levels in the cytoplasm of ductal cells within CP-like areas adjacent to the cancer cells (Figure 5, A-C). As in the normal pancreas, islet cells (outlined by arrowheads) did not exhibit Id immunoreactivity. In 4 of 9 CP samples, there were foci of ductal cell dysplasia of relatively large interlobular ducts, all of which exhibited moderate to strong Id-1, Id-2, and Id-3 immunoreactivity (Figure 5, D-F). Five of 9 CP samples also contained foci of large ducts exhibiting atypical papillary epithelium. Serial section analysis of one of those CP samples revealed mild to moderate Id-1 and Id-2 immunoreactivity and weak Id-3 immunoreactivity in the cells of these atypical papillary ducts (Figure 6, A-C). In contrast, in some of these CP samples, moderate to strong Id-3 immunoreactivity was also observed (Figure 6D). However, most of the ductal cells forming the typical ductular structures of CP, such as large interlobular ducts and small proliferating ducts, exhibited generally only weak to occasionally moderate Id immunoreactivity (data not shown).

The immunohistochemical data for Id-1, Id-2, and Id-3 are summarized in Table 1. In the case of Id-1 and Id-2, the cancer cells as well as the dysplastic and atypical papillary ducts in CP exhibited a significantly higher score than the ductal cells in the normal pancreas. In contrast, due to the marked variability in Id-3 immunostaining in the normal pancreas, the differences between normal and cancer cells and normal and dysplastic cells did not achieve statistical significance.

### Discussion

Id proteins constitute a family of HLH transcription factors that are important regulators of cellular differentiation and proliferation.<sup>1,2</sup> To date, four members of the human Id family have been identified.<sup>1,10-12</sup> Their expression is enhanced during cellular proliferation and in response to mitogenic stimuli,<sup>19,31</sup> and overexpression of Id genes inhibits differentiation and/or enhances proliferation in several different cell types.<sup>15,32-34</sup> The forced expression of Id-1 in mouse small intestinal epithelium results in



Table 1. Histological Scoring

		Id-1	Id-2	Id-3
Normal (n = 6)	Ductal cells	2.0 ± 0.4	2.3 ± 0.2	2.5 ± 0.9
Cancer (n = 10)	Cancer cells	4.5* ± 0.5	5.2 <sup>§</sup> ± 0.3	4.5 ± 0.6
CP (n = 9)	Typical CP lesions (n = 9)	2.7 ± 0.5	3.1 ± 0.6	3.4 ± 0.7
	Dysplastic ducts (n = 4)	5.3 <sup>†</sup> ± 0.2	5.8 <sup>†</sup> ± 0.2	5.3 ± 0.4
	Atypical papillary ducts (n = 5)	4.4 <sup>‡</sup> ± 0.2	5.2 <sup>‡</sup> ± 0.2	5.0 ± 0.4

Scoring of the histological specimens was performed as described in the Patients and Methods section. Values are the means ± SD of the number of samples indicated in parenthesis. *P* values are based on comparisons with the respective controls in the normal samples.

\**P* < 0.02; <sup>†</sup>*P* < 0.01; <sup>‡</sup>*P* = 0.004; <sup>§</sup>*P* = 0.001.

adenoma formation in these animals.<sup>35</sup> The growth-promoting effects of Id genes are thought to occur through several mechanisms. For example, Id-2 can bind to members of the pRB tumor suppressor family, thus blocking their growth-suppressing activity,<sup>20,21</sup> and Id-1 and Id-2 can antagonize the bHLH-mediated activation of known inhibitors of cell cycle progression such as the cyclin-dependent kinase inhibitor p21.<sup>23</sup>

In the present study, we determined by Northern blot analysis that a significant percentage of human pancreatic cancers expressed increased Id-1, Id-2, and Id-3 mRNA levels. Increased expression was most evident for Id-1 (6.5-fold) and Id-2 (fivefold). In contrast, Id-3 mRNA levels were only twofold increased in the cancer samples, partly because this mRNA was present at relatively high levels in the normal pancreas. Immunohistochemical analysis confirmed the presence of Id-1, Id-2, and Id-3 in the cancer cells within the tumor mass, whereas in the normal pancreas faint Id-1 and Id-2 immunoreactivity and moderate to occasionally strong Id-3 immunoreactivity was present in some ductal cells. Pancreatic acinar and islet cells in the normal pancreas were devoid of Id-1, Id-2, and Id-3 immunoreactivity. In the cancer samples, all three Id proteins often colocalized in the cancer cells. Coexpression of all three Id genes was also observed in cultured pancreatic cancer cell lines, which often exhibited a close correlation between Id mRNA and protein expression. However, in MIA-PaCa-2 there was a divergence of Id-2 mRNA and protein levels, and in PANC-1 cells, Id-3 mRNA levels did not correlate well with Id-3 protein expression. These observations suggest that in these cells, the half-life of either Id mRNA or Id protein may be altered by comparison with the other cell lines. Interestingly, Id-2 immunoblotting revealed two closely spaced bands of approximately 16 and 18 kd in 4 of 5 cell lines. In view of the fact that two possible initiation codons have been reported for the Id-2 gene,<sup>36</sup> our observation raises the possibility that the two Id-2-immunoreactive bands may represent separate translation products of the Id-2 gene.

Pancreatic cancers often harbor p53 tumor suppressor gene mutations<sup>37</sup> and exhibit alterations in apoptosis pathways. Thus, these cancers often exhibit increased expression of anti-apoptotic proteins such as Bcl-2<sup>38</sup> and abnormal resistance to Fas-ligand-mediated apoptosis.<sup>39</sup> It has been shown recently that forced constitutive expression of Id genes together with the expression of anti-apoptotic genes such as Bcl-2 or Bcl-X<sub>L</sub> can result in

malignant transformation of human fibroblasts,<sup>11</sup> raising the possibility that the enhanced Id expression in pancreatic cancers together with increased expression of anti-apoptotic genes may contribute to the malignant potential of pancreatic cancer cells *in vivo*.

In the CP tissues there was no significant increase in Id-1, Id-2, and Id-3 mRNA levels in comparison to the normal pancreas. Immunohistochemical analysis of pancreatic cancer samples revealed colocalization of weak to moderate Id-1, Id-2, and Id-3 immunoreactivity in proliferating ductal cells in the CP-like regions adjacent to the cancer cells, indicating that Id expression was not restricted to the cancer cells. Similarly, analysis of CP samples indicated weak Id-1, Id-2, and Id-3 immunoreactivity in the cells of small proliferating ducts and large ducts without dysplastic changes. In general, there was a correlation between weak immunoreactivity and low Id mRNA levels. However, in samples that harbored large ducts with papillary structures there was moderate Id immunoreactivity, and in the cells forming dysplastic ducts there was moderate to strong Id immunoreactivity. In these CP samples, Id mRNA levels were relatively higher than in the CP samples that were devoid of these histological changes. Overall, however, increased Id expression, most notably of Id-1 and Id-2, distinguished a subgroup of pancreatic cancers from CP (Table 1).

Epidemiological studies have shown that the risk of developing pancreatic cancer is increased up to 16-fold in patients with pre-existing CP in comparison to the general population.<sup>40</sup> The mechanisms that contribute to neoplastic transformation in CP are not known. Although there is no established tumor progression model for pancreatic cancer, such as the adenoma-carcinoma sequence of colorectal carcinoma,<sup>41</sup> it is generally accepted that K-ras and p16 mutations occur relatively early in pancreatic carcinogenesis, whereas p53 mutations occur late in this process.<sup>37,41-43</sup> Increased Id expression may contribute to malignant transformation of cultured cell lines *in vitro*<sup>11</sup> and has been linked to cell invasion in a murine mammary epithelial cell line.<sup>44</sup> In view of the current findings that Id-1, Id-2, and Id-3 are overexpressed in pancreatic cancer and in dysplastic/metaplastic ducts in CP, these observations raise the possibility that elevated levels of Id-1, Id-2, and, to a lesser extent, Id-3 may represent relatively early markers of pancreatic malignant transformation and may contribute to the pathobiology of pancreatic cancer.

## References

- Jan YN, Jan LY: HLH proteins, fly neurogenesis, and vertebrate myogenesis. *Cell* 1993, 75:827-830
- Olson EN, Klein WH: bHLH factors in muscle development: dead lines and commitments, what to leave in and what to leave out. *Genes Dev* 1994, 8:1-8
- Begley CG, Aplan PD, Denning SM, Haynes BF, Waldmann TA, Kirsch IR: The gene SCL is expressed during early hematopoiesis and encodes a differentiation-related DNA-binding motif. *Proc Natl Acad Sci USA* 1989, 86:10128-10132
- Johnson JE, Birren SJ, Anderson DJ: Two rat homologues of *Drosophila achaete-scute* specifically expressed in neuronal precursors. *Nature* 1990, 346:858-861
- Weintraub H: The MyoD family and myogenesis: redundancy, networks, and thresholds. *Cell* 1993, 75:1241-1244
- Hu JS, Olson EN, Kingston RE: HEB, a helix-loop-helix protein related to E2A, and Irf2 that can modulate the DNA-binding ability of myogenic regulatory factors. *Mol Cell Biol* 1992, 12:1031-1042
- Langlands K, Yin X, Anand G, Prochownik EV: Differential interactions of Id proteins with basic-helix-loop-helix transcription factors. *J Biol Chem* 1997, 272:19785-19793
- Murre C, Bain G, van Dijk MA, Engel I, Furnari BA, Massari ME, Matthews JR, Quong MW, Rivera RR, Stuver MH: Structure and function of helix-loop-helix proteins. *Biochim Biophys Acta* 1994, 1218:129-135
- Murre C, McCaw PS, Vaessin H, Caudy M, Jan LY, Jan YN, Cabrera CV, Buskin JN, Hauschka SD, Lassar AB, Baltimore D: Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* 1989, 58:537-544
- Benezra R, Davis RL, Lockshon D, Turner DL, Weintraub H: The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. *Cell* 1990, 61:49-59
- Norton JD, Atherton GT: Coupling of cell growth control and apoptosis functions of Id proteins. *Mol Cell Biol* 1998, 18:2371-2381
- Norton JD, Deed RW, Craggs G, Sablitzky F: Id helix-loop-helix proteins in cell growth and differentiation. *Trends Cell Biol* 1998, 8:58-65
- Christy BA, Sanders LK, Lau LF, Copeland NG, Jenkins NA, Nathans D: An Id-related helix-loop-helix protein encoded by a growth factor-inducible gene. *Proc Natl Acad Sci USA* 1991, 88:1815-1819
- Kawaguchi N, DeLuca HF, Noda M: Id gene expression and its suppression by 1,25-dihydroxyvitamin D3 in rat osteoblastic osteosarcoma cells. *Proc Natl Acad Sci USA* 1992, 89:4569-4572
- Kreider BL, Benezra R, Rovera G, Kadesch T: Inhibition of myeloid differentiation by the helix-loop-helix protein Id. *Science* 1992, 255:1700-1702
- Le Jossic C, Ilyin GP, Loyer P, Glaire D, Cariou S, Guguen-Guillouzo C: Expression of helix-loop-helix factor Id-1 is dependent on the hepatocyte proliferation and differentiation status in rat liver and in primary culture. *Cancer Res* 1994, 54:6065-6068
- Sun XH, Copeland NG, Jenkins NA, Baltimore D: Id proteins Id1 and Id2 selectively inhibit DNA binding by one class of helix-loop-helix proteins. *Mol Cell Biol* 1991, 11:5603-5611
- Wilson RB, Kiledjian M, Shen CP, Benezra R, Zwollo P, Dymecki SM, Desiderio SV, Kadesch T: Repression of immunoglobulin enhancers by the helix-loop-helix protein Id: implications for B-lymphoid cell development. *Mol Cell Biol* 1991, 11:6185-6191
- Hara E, Yamaguchi T, Nojima H, Ide T, Campisi J, Okayama H, Oda K: Id-related genes encoding helix-loop-helix proteins are required for G1 progression and are repressed in senescent human fibroblasts. *J Biol Chem* 1994, 269:2139-2145
- Iavarone A, Garg P, Lasorella A, Hsu J, Israel MA: The helix-loop-helix protein Id-2 enhances cell proliferation and binds to the retinoblastoma protein. *Genes Dev* 1994, 8:1270-1284
- Lasorella A, Iavarone A, Israel MA: Id2 specifically alters regulation of the cell cycle by tumor suppressor proteins. *Mol Cell Biol* 1996, 16:2570-2578
- Peverali FA, Ramqvist T, Saffrich R, Pepperkok R, Barone MV, Philipson L: Regulation of G1 progression by E2A and Id helix-loop-helix proteins. *EMBO J* 1994, 13:4291-4301
- Prabhu S, Ignatova A, Park ST, Sun XH: Regulation of the expression of cyclin-dependent kinase inhibitor p21 by E2A and Id proteins. *Mol Cell Biol* 1997, 17:5888-5896
- Warshaw AL, Fernandez-del Castillo C: Pancreatic carcinoma. *N Engl J Med* 1992, 326:455-465
- Korc M: Role of growth factors in pancreatic cancer. *Surg Oncol Clin North Am* 1998, 7:25-41
- Kleeff J, Ishiwata T, Friess H, Büchler MW, Israel MA, Korc M: The helix-loop-helix protein Id2 is overexpressed in human pancreatic cancer. *Cancer Res* 1998, 58:3769-3772
- Oertel JE, Heffes CS, Oertel YC: *Pancreas. Diagnostic Surgical Pathology*. Edited by SS Sternberg. New York, Raven Press, 1989, pp. 1057-1093
- Korc M, Chandrasekar B, Yamanaka Y, Friess H, Büchler MW, Beger HG: Overexpression of the epidermal growth factor receptor in human pancreatic cancer is associated with concomitant increase in the levels of epidermal growth factor and transforming growth factor  $\alpha$ . *J Clin Invest* 1992, 90:1352-1360
- Saeki T, Stromberg K, Qi CF, Gullick WJ, Tahara E, Normanno N, Ciardiello F, Kenney N, Johnson GR, Salomon DS: Differential immunohistochemical detection of amphiregulin and cripto in human normal colon and colorectal tumors. *Cancer Res* 1992, 52:3467-3473
- Cantero D, Friess H, Delforin J, Zimmermann A, Bründler MA, Riesle E, Korc M, Büchler MW: Enhanced expression of urokinase plasminogen activator and its receptor in pancreatic carcinoma. *Br J Cancer* 1997, 75:388-395
- Desprez PY, Hara E, Bissell MJ, Campisi J: Suppression of mammary epithelial cell differentiation by the helix-loop-helix protein Id-1. *Mol Cell Biol* 1995, 15:3398-3404
- Shoji W, Yamamoto T, Obinata M: The helix-loop-helix protein Id inhibits differentiation of murine erythroleukemia cells. *J Biol Chem* 1994, 269:5078-5084
- Cross JC, Flannery ML, Blonar MA, Steingrimsson E, Jenkins NA, Copeland NG, Rutter WJ, Werb Z: Hxl encodes a basic helix-loop-helix transcription factor that regulates trophoblast cell development. *Development* 1995, 121:2513-2523
- Sun XH: Constitutive expression of the Id1 gene impairs mouse B cell development. *Cell* 1994, 79:893-900
- Wice BM, Gordon JI: Forced expression of Id-1 in the adult mouse small intestinal epithelium is associated with development of adenomas. *J Biol Chem* 1998, 273:25310-25319
- Barone MV, Pepperkok R, Peverali FA, Philipson L: Id proteins control growth induction in mammalian cells. *Proc Natl Acad Sci USA* 1994, 91:4985-4988
- Barton CM, Staddon SL, Hughes CM, Hall PA, O'Sullivan C, Kloppel G, Theis B, Russell RC, Neoptolimos J, Williamson RCN, Lane DP, Lemoine NR: Abnormalities of the p53 tumour suppressor gene in human pancreatic cancer. *Br J Cancer* 1991, 64:1076-1082
- Ohshio G, Suwa H, Imamura T, Yamaki K, Tanaka T, Hashimoto Y, Imamura M: An immunohistochemical study of bcl-2 and p53 protein expression in pancreatic carcinomas. *Scand J Gastroenterol* 1998, 33:535-539
- Ungeloren H, Voss M, Jansen M, Roeder C, Henne-Bruns D, Kremer B, Kalthoff H: Human pancreatic adenocarcinomas express Fas and Fas ligand yet are resistant to Fas-mediated apoptosis. *Cancer Res* 1998, 58:1741-1749
- Niedermaier C, Niedermaier MC, Heintges T, Lüthen R: Epidemiology: relation between chronic pancreatitis and pancreatic carcinoma. *Cancer of the Pancreas*. Edited by HG Beger, MW Büchler, MH Schoenberg. Ulm, Germany: Universitätsverlag Ulm GmbH, 1996, pp 6-9
- Moskaluk CA, Kern SE: Molecular genetics of pancreatic carcinoma. *Pancreatic Cancer: Pathogenesis, Diagnosis, and Treatment*. Edited by HA Reber. Totowa, NJ, Humana Press, 1998, pp 3-20
- Moskaluk CA, Hruban RH, Kern SE: p16 and K-ras gene mutations in the intraductal precursors of human pancreatic adenocarcinoma. *Cancer Res* 1997, 57:2140-2143
- Tada M, Ohashi M, Shiratori Y, Okudaira T, Komatsu Y, Kawabe T, Yoshida H, Machinami R, Kishi K, Ornata M: Analysis of K-ras gene mutation in hyperplastic duct cells of the pancreas without pancreatic disease. *Gastroenterology* 1996, 110:227-231
- Desprez PY, Lin CO, Thomasset N, Simpson CJ, Bissell MJ, Campisi J: A novel pathway for mammary epithelial cell invasion induced by the helix-loop-helix protein Id-1. *Mol Cell Biol* 1998, 18:4577-4588



**Tightly regulated and inducible expression of a yoked hormone-receptor complex in HEK 293 cells.**

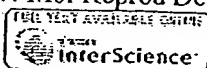
Meehan TP, Puett D, Narayan P.

Department of Biochemistry and Molecular Biology, University of Georgia, Athens, Georgia 30602, USA.

We have previously reported the construction of a constitutively active luteinizing hormone receptor by covalently linking a fused heterodimeric hormone to the extracellular domain of the G protein-coupled receptor. This yoked hormone-receptor complex (YHR) was found to produce high levels of cAMP in the absence of exogenous hormone. Stable lines expressing YHR were generated in HEK 293 cells to obtain lines with different expression levels; however, in a relatively short time of continued passage, it was found that YHR expression was greatly reduced. Herein, we describe the development of clonal lines of HEK 293 cells in which the expression of YHR is under the control of a tetracycline-regulated system. Characterization of clonal lines revealed tight control of YHR expression both by dose and time of incubation with doxycycline. These experiments demonstrated a good correlation between expression levels of the receptor and basal cAMP production. Moreover, the reduction in receptor expression following doxycycline removal revealed that YHR mRNA and protein decayed at similar rates, again suggesting a strong linkage between mRNA and protein levels. The controlled expression of YHR in this cell system will allow for a more detailed analysis of the signaling properties associated with constitutive receptor activation and may prove to be advantageous in developmental studies with transgenic animals.

PMID: 14766006 [PubMed - indexed for MEDLINE]





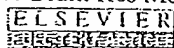
**c-fos and estrogen receptor gene expression pattern in the rat uterine epithelium during the estrous cycle.**

Mendoza-Rodriguez CA, Merchant-Larios H, Segura-Valdez ML, Moreno-Mendoza N, Cruz ME, Arteaga-Lopez P, Camacho-Arroyo I, Dominguez R, Cerbon M.

Facultad de Quimica, Universidad Nacional Autonoma de Mexico, Ciudad Universitaria, Coyoacan 04510, Mexico, D.F., Mexico.

Different studies in ovariectomized estrogen treated animals support the idea that c-fos plays a role in the proliferation of uterine epithelial cells. However, these studies invite us to reassess the role played by c-fos in epithelial cell types of the endometrium during the estrous cycle. The present study was undertaken to determine the c-fos and estrogen receptor (ER) gene expression pattern in the rat uterine epithelium during the estrous cycle in which natural and cyclic changes of steroid hormones occur, and correlate these changes with the proliferation status of this cellular types. Proliferation was assessed during the estrous cycle using bromodeoxyuridine incorporation to DNA. ERalpha and beta proteins were assessed by immunohistochemistry. The regulation of c-fos gene expression in the uterus of intact animals during the estrous cycle was evaluated using both in situ hybridization and immunohistochemistry. Estradiol (E(2)) and progesterone (P(4)) plasma levels were assessed by radioimmunoassay. The results indicated that luminal (LE) and glandular epithelia (GE) presented maximal proliferation during the metestrus (M) and the diestrus (D) days. However, during the proestrus (P) day only LE presented proliferation, and during the estrus (E) day only the stromal cells proliferated. A marked immunostaining for ERalpha was detected in both LE and GE cells during the early phases of the cycle but diminished on the P and the E day. In contrast, ERbeta was undetectable in both epithelia during all stages of the cycle. The highest c-fos mRNA level was detected in both epithelia on the M day, followed by a significant reduction during the other days of the cycle. The highest protein content was observed on the M and D days, and the minimal value was detected on the E day. The c-Fos protein level in LE was increased during M and D days, presenting a high correlation with the cellular proliferation pattern of this cell type. In conclusion, the overall results indicate that c-Fos protein presented a good correlation with uterine epithelial cell proliferation of LE. In the case of GE, the same tendency was observed, although no significant correlation was found. Both in LE and GE, c-fos mRNA did not strictly correlate with its protein levels. c-fos seems to have a postranscriptional regulation in uterine epithelial cells during the rat's estrous cycle. Copyright 2003 Wiley-Liss, Inc.

PMID: 12589649 [PubMed - indexed for MEDLINE]



**[3H]MK-801 binding and the mRNA for the NMDAR1 subunit of the NMDA receptor are differentially distributed in human and rat forebrain.**

**Meoni P, Mugnaini M, Bunnemann BH, Trist DG, Bowery NG.**

Department of Pharmacology, Medical School, University of Birmingham, UK.  
meonip@novel15.bham.ac.uk

The distributions of [3H]MK-801 binding and the NMDA NR1 subunit mRNA were studied using receptor autoradiography and in-situ hybridization in rat and human brain whole-hemisphere coronal sections. Receptor protein detected by radioligand autoradiography and the mRNA for the key subunit of the receptor presented similar distributions in the forebrain, with a few areas showing an imbalance between the levels of mRNA and receptor protein. Human frontal cortex showed a relative abundance of NMDAR1 mRNA as compared to [3H]MK-801 binding. The same area in rat brain did not show any difference in the two distributions. In comparison, the rat claustrum presented a relative excess of NMDAR1 mRNA which was not detected in human sections. Human caudate nucleus exhibited relatively high levels of [3H]MK-801 binding that were unmatched in rat caudate. The hippocampi of either species presented similar levels of [3H]MK-801 binding and NMDAR1 mRNA, but when the two signals were measured in specific subfields of the hippocampal formation, the differential distribution of the two signals reflected the anatomy of hippocampal connections assuming a preferential dendritic distribution for MK-801 binding. Interestingly, rat and human hippocampi also showed some important species-dependent difference in the relative distribution of the receptor protein and mRNA. The data presented show an overall good correlation between the mRNA for the key subunit of the NMDA receptor and the functional receptor detected with radioligand binding and highlight the presence of local differences in their ratio. This may reflect different splicing of the mRNA for the NMDAR1 subunit in specific brain areas of rat and human. The species-dependent differences in the relative distribution of the mRNA for the key subunit of the NMDA receptor and that of a marker of functional receptors also highlights important differences in the NMDA function in rat and human brain.

PMID: 9526033 [PubMed - indexed for MEDLINE]

## Review

# Translation Initiation in Cancer: A Novel Target for Therapy<sup>1</sup>

Funda Meric<sup>2</sup> and Kelly K. Hunt

Department of Surgical Oncology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030

### Abstract

Translation initiation is regulated in response to nutrient availability and mitogenic stimulation and is coupled with cell cycle progression and cell growth. Several alterations in translational control occur in cancer. Variant mRNA sequences can alter the translational efficiency of individual mRNA molecules, which in turn play a role in cancer biology. Changes in the expression or availability of components of the translational machinery and in the activation of translation through signal transduction pathways can lead to more global changes, such as an increase in the overall rate of protein synthesis and translational activation of the mRNA molecules involved in cell growth and proliferation. We review the basic principles of translational control, the alterations encountered in cancer, and selected therapies targeting translation initiation to help elucidate new therapeutic avenues.

### Introduction

The fundamental principle of molecular therapeutics in cancer is to exploit the differences in gene expression between cancer cells and normal cells. With the advent of cDNA array technology, most efforts have concentrated on identifying differences in gene expression at the level of mRNA, which can be attributable either to DNA amplification or to differences in transcription. Gene expression is quite complicated, however, and is also regulated at the level of mRNA stability, mRNA translation, and protein stability.

The power of translational regulation has been best recognized among developmental biologists, because transcription does not occur in early embryogenesis in eukaryotes. For example, in *Xenopus*, the period of transcriptional quiescence continues until the embryo reaches midblastula transition, the 4000-cell stage. Therefore, all necessary mRNA molecules are transcribed during oogenesis and stockpiled in a translationally inactive, masked form. The mRNA are translationally activated at appropriate times during oocyte maturation, fertilization, and

early embryogenesis and thus, are under strict translational control.

Translation has an established role in cell growth. Basically, an increase in protein synthesis occurs as a consequence of mitogenesis. Until recently, however, little was known about the alterations in mRNA translation in cancer, and much is yet to be discovered about their role in the development and progression of cancer. Here we review the basic principles of translational control, the alterations encountered in cancer, and selected therapies targeting translation initiation to elucidate potential new therapeutic avenues.

### Basic Principles of Translational Control

#### Mechanism of Translation Initiation

Translation initiation is the main step in translational regulation. Translation initiation is a complex process in which the initiator tRNA and the 40S and 60S ribosomal subunits are recruited to the 5' end of a mRNA molecule and assembled by eukaryotic translation initiation factors into an 80S ribosome at the start codon of the mRNA (Fig. 1). The 5' end of eukaryotic mRNA is capped, i.e., contains the cap structure m<sup>7</sup>GpppN (7-methyl-guanosine-triphospho-5'-ribonucleoside). Most translation in eukaryotes occurs in a cap-dependent fashion, i.e., the cap is specifically recognized by the eIF4E,<sup>3</sup> which binds the 5' cap. The eIF4F translation initiation complex is then formed by the assembly of eIF4E, the RNA helicase eIF4A, and eIF4G, a scaffolding protein that mediates the binding of the 40S ribosomal subunit to the mRNA molecule through interaction with the eIF3 protein present on the 40S ribosome. eIF4A and eIF4B participate in melting the secondary structure of the 5' UTR of the mRNA. The 43S initiation complex (40S/eIF2/Met-tRNA/GTP complex) scans the mRNA in a 5'→3' direction until it encounters an AUG start codon. This start codon is then base-paired to the anticodon of initiator tRNA, forming the 48S initiation complex. The initiation factors are then displaced from the 48S complex, and the 60S ribosome joins to form the 80S ribosome.

Unlike most eukaryotic translation, translation initiation of certain mRNAs, such as the picornavirus RNA, is cap independent and occurs by internal ribosome entry. This mechanism does not require eIF4E. Either the 43S complex can bind the initiation codon directly through interaction with the IRES in the 5' UTR such as in the encephalomyocarditis virus, or it can

Received 5/16/02; revised 7/12/02; accepted 7/22/02.

<sup>1</sup>F. M. is supported by The University of Texas M. D. Anderson Cancer Center Physician-Scientist Program and by NIH Grant 1K08-CA 91895-01. K. K. H. is supported by Department of Defense Award DAMD-17-97-1-7162.

<sup>2</sup>To whom requests for reprints should be addressed, at Department of Surgical Oncology, Box 444, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030. Phone: (713) 745-4453; Fax: (713) 745-4926; E-mail: fmeric@mdanderson.org.

<sup>3</sup>The abbreviations used are: eIF4E, eukaryotic initiation factor 4E; UTR, untranslated region; IRES, internal ribosome entry site; 4E-BP1, eukaryotic initiation factor 4E-binding protein 1; S6K, ribosomal p70 S6 kinase; mTOR, mammalian target of rapamycin; ATM, ataxia telangiectasia mutated; PI3K, phosphatidylinositol 3-kinase; PTEN, phosphatase and tensin homolog deleted from chromosome 10; PP2A, protein phosphatase 2A; TGF- $\beta$ 3, transforming growth factor- $\beta$ 3; PAP, poly(A) polymerase; EPA, eicosapentaenoic acid; mda-7, melanoma differentiation-associated gene 7.

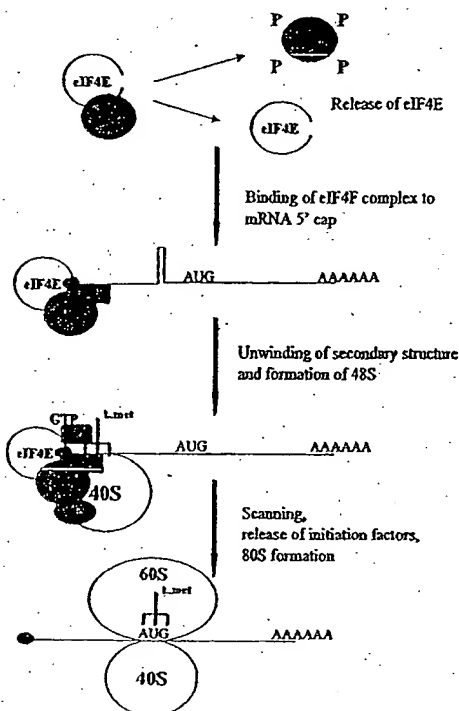


Fig. 1. Translation initiation in eukaryotes. The 4E-BPs are hyperphosphorylated to release eIF4E so that it can interact with the 5' cap, and the eIF4F initiation complex is assembled. The interaction of poly(A) binding protein with the initiation complex and circularization of the mRNA is not depicted in the diagram. The secondary structure of the 5' UTR is melted, the 40S ribosomal subunit is bound to eIF3, and the ternary complex consisting of eIF2, GTP, and the Met-tRNA are recruited to the mRNA. The ribosome scans the mRNA in a 5'→3' direction until an AUG start codon is found in the appropriate sequence context. The initiation factors are released, and the large ribosomal subunit is recruited.

initially attach to the IRES and then reach the initiation codon by scanning or transfer, as is the case with the poliovirus (1).

#### Regulation of Translation Initiation

Translation initiation can be regulated by alterations in the expression or phosphorylation status of the various factors involved. Key components in translational regulation that may provide potential therapeutic targets follow.

**eIF4E.** eIF4E plays a central role in translation regulation. It is the least abundant of the initiation factors and is considered the rate-limiting component for initiation of cap-dependent translation. eIF4E may also be involved in mRNA splicing, mRNA 3' processing, and mRNA nucleocytoplasmic transport (2). eIF4E expression can be increased at the transcriptional level in response to serum or growth factors (3). eIF4E overexpression may cause preferential translation of mRNAs containing excessive secondary structure in their 5' UTR that are normally discriminated against by the trans-

lational machinery and thus are inefficiently translated (4–7). As examples of this, overexpression of eIF4E promotes increased translation of vascular endothelial growth factor, fibroblast growth factor-2, and cyclin D1 (2, 8, 9).

Another mechanism of control is the regulation of eIF4E phosphorylation. eIF4E phosphorylation is mediated by the mitogen-activated protein kinase-interacting kinase 1, which is activated by the mitogen-activated pathway, activating extracellular signal-related kinases and the stress-activated pathway acting through p38 mitogen-activated protein kinase (10–13). Several mitogens, such as serum, platelet-derived growth factor, epidermal growth factor, insulin, angiotensin II, src kinase overexpression, and ras overexpression, lead to eIF4E phosphorylation (14). The phosphorylation status of eIF4E is usually correlated with the translational rate and growth status of the cell; however, eIF4E phosphorylation has also been observed in response to some cellular stresses when translational rates actually decrease (15). Thus, further study is needed to understand the effects of eIF4E phosphorylation on eIF4E activity.

Another mechanism of regulation is the alteration of eIF4E availability by the binding of eIF4E to the eIF4E-binding proteins (4E-BPs, also known as PHAS-I). 4E-BPs compete with eIF4G for a binding site in eIF4E. The binding of eIF4E to the best characterized eIF4E-binding protein, 4E-BP1, is regulated by 4E-BP1 phosphorylation. Hypophosphorylated 4E-BP1 binds to eIF4E, whereas 4E-BP1 hyperphosphorylation decreases this binding. Insulin, angiotensin, epidermal growth factor, platelet-derived growth factor, hepatocyte growth factor, nerve growth factor, insulin-like growth factors I and II, interleukin 3, granulocyte-macrophage colony-stimulating factor + steel factor, gastrin, and the adenovirus have all been reported to induce phosphorylation of 4E-BP1 and to decrease the ability of 4E-BP1 to bind eIF4E (15, 16). Conversely, deprivation of nutrients or growth factors results in 4E-BP1 dephosphorylation, an increase in eIF4E binding, and a decrease in cap-dependent translation.

**p70 S6 Kinase.** Phosphorylation of ribosomal 40S protein S6 by S6K is thought to play an important role in translational regulation. S6K  $-/-$  mouse embryonic cells proliferate more slowly than do parental cells, demonstrating that S6K has a positive influence on cell proliferation (17). S6K regulates the translation of a group of mRNAs possessing a 5' terminal oligopyrimidine tract (5' TOP) found at the 5' UTR of ribosomal protein mRNAs and other mRNAs coding for components of the translational machinery. Phosphorylation of S6K is regulated in part based on the availability of nutrients (18, 19) and is stimulated by several growth factors, such as platelet-derived growth factor and insulin-like growth factor I (20).

**eIF2 $\alpha$  Phosphorylation.** The binding of the initiator tRNA to the small ribosomal unit is mediated by translation initiation factor eIF2. Phosphorylation of the  $\alpha$ -subunit of eIF2 prevents formation of the eIF2/GTP/Met-tRNA complex and inhibits global protein synthesis (21, 22). eIF2 $\alpha$  is phosphorylated under a variety of conditions, such as viral infection, nutrient deprivation, heme deprivation, and apoptosis (22). eIF2 $\alpha$  is phosphorylated by heme-regulated inhibitor, nutrient-regulated protein kinase, and the IFN-induced, double-stranded RNA-activated protein kinase (PKR; Ref. 23).

**The mTOR Signaling Pathway.** The macrolide antibiotic rapamycin (Sirinrus; Wyeth-Ayerst Research, Collegeville, PA) has been the subject of intensive study because it inhibits signal transduction pathways involved in T-cell activation. The rapamycin-sensitive component of these pathways is mTOR (also called FRAP or RAFT1). mTOR is the mammalian homologue of the yeast TOR proteins that regulate  $G_1$  progression and translation in response to nutrient availability (24). mTOR is a serine-threonine kinase that modulates translation initiation by altering the phosphorylation status of 4E-BP1 and S6K (Fig. 2; Ref. 25).

4E-BP1 is phosphorylated on multiple residues. mTOR phosphorylates the Thr-37 and Thr-46 residues of 4E-BP1 *in vitro* (26); however, phosphorylation at these sites is not associated with a loss of eIF4E binding. Phosphorylation of Thr-37 and Thr-46 is required for subsequent phosphorylation at several COOH-terminal, serum-sensitive sites; a combination of these phosphorylation events appears to be needed to inhibit the binding of 4E-BP1 to eIF4E (25). The product of the ATM gene, p38/MSK1 pathway, and protein kinase C $\alpha$  also play a role in 4E-BP1 phosphorylation (27-29).

S6K and 4E-BP1 are also regulated, in part, by PI3K and its downstream protein kinase Akt. PTEN is a phosphatase that negatively regulates PI3K signaling. PTEN null cells have constitutively active Akt, with increased S6K activity and S6 phosphorylation (30). S6K activity is inhibited both by PI3K inhibitors wortmannin and LY294002 and by mTOR inhibitor rapamycin (24). Akt phosphorylates Ser-2448 in mTOR *in vitro*, and this site is phosphorylated upon Akt activation *in vivo* (31-33). Thus, mTOR is regulated by the PI3K/Akt pathway; however, this does not appear to be the only mode of regulation of mTOR activity. Whether the PI3K pathway also regulates S6K and 4E-BP1 phosphorylation independent of mTOR is controversial.

Interestingly, mTOR autophosphorylation is blocked by wortmannin but not by rapamycin (34). This seeming inconsistency suggests that mTOR-responsive regulation of 4E-BP1 and S6K activity occurs through a mechanism other than intrinsic mTOR kinase activity. An alternate pathway for 4E-BP1 and S6K phosphorylation by mTOR activity is by the inhibition of a phosphatase. Treatment with calyculin A, an inhibitor of phosphatases 1 and 2A, reduces rapamycin-induced dephosphorylation of 4E-BP1 and S6K by rapamycin (35). PP2A interacts with full-length S6K but not with a S6K mutant that is resistant to dephosphorylation resulting from rapamycin. mTOR phosphorylates PP2A *in vitro*; however, how this process alters PP2A activity is not known. These results are consistent with the model that phosphorylation of a phosphatase by mTOR prevents dephosphorylation of 4E-BP1 and S6K, and conversely, that nutrient deprivation and rapamycin block inhibition of the phosphatase by mTOR.

**Polyadenylation.** The poly(A) tail in eukaryotic mRNA is important in enhancing translation initiation and mRNA stability. Polyadenylation plays a key role in regulating gene expression during oogenesis and early embryogenesis. Some mRNA that are translationally inactive in the oocyte are polyadenylated concomitantly with translational activation in oocyte maturation, whereas other mRNAs that are translationally active during oogenesis are deadenylated and trans-

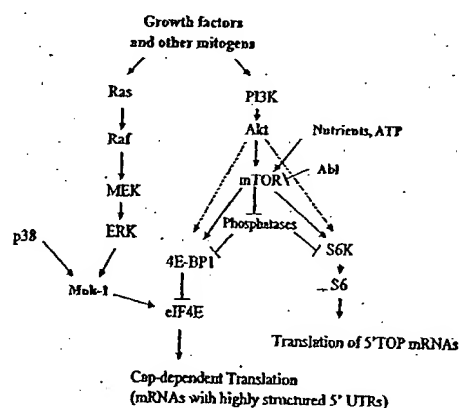


Fig. 2. Regulation of translation initiation by signal transduction pathways. Signaling via p38, extracellular signal-related kinase, PI3K, and mTOR can all activate translation initiation.

lationally silenced (36-38). Thus, control of poly(A) tail synthesis is an important regulatory step in gene expression. The 5' cap and poly(A) tail are thought to function synergistically to regulate mRNA translational efficiency (39, 40).

**RNA Packaging.** Most RNA-binding proteins are assembled on a transcript at the time of transcription, thus determining the translational fate of the transcript (41). A highly conserved family of Y-box proteins is found in cytoplasmic messenger ribonucleoprotein particles, where the proteins are thought to play a role in restricting the recruitment of mRNA to the translational machinery (41-43). The major mRNA-associated protein, YB-1, destabilizes the interaction of eIF4E and the 5' mRNA cap *in vitro*, and overexpression of YB-1 results in translational repression *in vivo* (44). Thus, alterations in RNA packaging can also play an important role in translational regulation.

#### Translation Alterations Encountered in Cancer

Three main alterations at the translational level occur in cancer: variations in mRNA sequences that increase or decrease translational efficiency, changes in the expression or availability of components of the translational machinery, and activation of translation through aberrantly activated signal transduction pathways. The first alteration affects the translation of an individual mRNA that may play a role in carcinogenesis. The second and third alterations can lead to more global changes, such as an increase in the overall rate of protein synthesis, and the translational activation of several mRNA species.

#### Variations in mRNA Sequence

Variations in mRNA sequence affect the translational efficiency of the transcript. A brief description of these variations and examples of each mechanism follow.

**Mutations.** Mutations in the mRNA sequence, especially in the 5' UTR, can alter its translational efficiency, as seen in the following examples.

*c-myc*. Saito *et al.* proposed that translation of full-length *c-myc* is repressed, whereas in several Burkitt lymphomas that have deletions of the mRNA 5' UTR, translation of *c-myc* is more efficient (45). More recently, it was reported that the 5' UTR of *c-myc* contains an IRES, and thus *c-myc* translation can be initiated by a cap-independent as well as a cap-dependent mechanism (46, 47). In patients with multiple myeloma, a C→T mutation in the *c-myc* IRES was identified (48) and found to cause an enhanced initiation of translation via internal ribosomal entry (49).

**BRCA1.** A somatic point mutation (117 G→C) in position -3 with respect to the start codon of the *BRCA1* gene was identified in a highly aggressive sporadic breast cancer (50). Chimeric constructs consisting of the wild-type or mutated *BRCA1* 5' UTR and a downstream luciferase reporter demonstrated a decrease in the translational efficiency with the 5' UTR mutation.

**Cyclin-dependent Kinase Inhibitor 2A.** Some inherited melanoma kindreds have a G→T transversion at base -34 of cyclin-dependent kinase inhibitor-2A, which encodes a cyclin-dependent kinase 4/cyclin-dependent kinase 6 kinase inhibitor important in G<sub>1</sub> checkpoint regulation (51). This mutation gives rise to a novel AUG translation initiation codon, creating an upstream open reading frame that competes for scanning ribosomes and decreases translation from the wild-type AUG.

**Alternate Splicing and Alternate Transcription Start Sites.** Alterations in splicing and alternate transcription sites can lead to variations in 5' UTR sequence, length, and secondary structure, ultimately impacting translational efficiency.

**ATM.** The *ATM* gene has four noncoding exons in its 5' UTR that undergo extensive alternative splicing (52). The contents of 12 different 5' UTRs that show considerable diversity in length and sequence have been identified. These divergent 5' leader sequences play an important role in the translational regulation of the *ATM* gene.

**mdm.** In a subset of tumors, overexpression of the oncoprotein *mdm2* results in enhanced translation of the *mdm2* mRNA. Use of different promoters leads to two *mdm2* transcripts that differ only in their 5' leaders (53). The longer 5' UTR contains two upstream open reading frames, and this mRNA is loaded with ribosomes inefficiently compared with the short 5' UTR.

**BRCA1.** In a normal mammary gland, *BRCA1* mRNA is expressed with a shorter leader sequence (5' UTRa), whereas in sporadic breast cancer tissue, *BRCA1* mRNA is expressed with a longer leader sequence (5' UTRb); the translational efficiency of transcripts containing 5' UTRb is 10 times lower than that of transcripts containing 5' UTRa (54).

**TGF-β3.** *TGF-β3* mRNA includes a 1.1-kb 5' UTR, which exerts an inhibitory effect on translation. Many human breast cancer cell lines contain a novel *TGF-β3* transcript with a 5' UTR that is 870 nucleotides shorter and has a 7-fold greater translational efficiency than the normal *TGF-β3* mRNA (55).

**Alternate Polyadenylation Sites.** Multiple polyadenylation signals leading to the generation of several transcripts with differing 3' UTR have been described for several mRNA species, such as the *RET* proto-oncogene (56), *ATM* gene (52), tissue inhibitor of metalloproteinases-3 (57), *RHOA*

proto-oncogene (58), and calmodulin-1 (59). Although the effect of these alternate 3' UTRs on translation is not yet known, they may be important in RNA-protein interactions that affect translational recruitment. The role of these alterations in cancer development and progression is unknown.

### Alterations in the Components of the Translation Machinery

Alterations in the components of translation machinery can take many forms.

**Overexpression of eIF4E.** Overexpression of eIF4E causes malignant transformation in rodent cells (60) and the deregulation of HeLa cell growth (61). Polunovsky *et al.* (62) found that eIF4E overexpression substitutes for serum and individual growth factors in preserving viability of fibroblasts, which suggests that eIF4E can mediate both proliferative and survival signaling.

Elevated levels of eIF4E mRNA have been found in a broad spectrum of transformed cell lines (63). eIF4E levels are elevated in all ductal carcinoma *in situ* specimens and invasive ductal carcinomas, compared with benign breast specimens evaluated with Western blot analysis (64, 65). Preliminary studies suggest that this overexpression is attributable to gene amplification (66).

There are accumulating data suggesting that eIF4E overexpression can be valuable as a prognostic marker. eIF4E overexpression was found in a retrospective study to be a marker of poor prognosis in stages I to III breast carcinoma (67). Verification of the prognostic value of eIF4E in breast cancer is now under way in a prospective trial (67). However, in a different study, eIF4E expression was correlated with the aggressive behavior of non-Hodgkin's lymphomas (68). In a prospective analysis of patients with head and neck cancer, elevated levels of eIF4E in histologically tumor-free surgical margins predicted a significantly increased risk of local-regional recurrence (9). These results all suggest that eIF4E overexpression can be used to select patients who might benefit from more aggressive systemic therapy. Furthermore, the head and neck cancer data suggest that eIF4E overexpression is a field defect and can be used to guide local therapy.

**Alterations in Other Initiation Factors.** Alterations in a number of other initiation factors have been associated with cancer. Overproduction of eIF4G, similar to eIF4E, leads to malignant transformation *in vitro* (69). eIF-2α is found in increased levels in bronchioloalveolar carcinomas of the lung (3). Initiation factor eIF-4A1 is overexpressed in melanoma (70) and hepatocellular carcinoma (71). The p40 subunit of translation initiation factor 3 is amplified and overexpressed in breast and prostate cancer (72), and the eIF3-p110 subunit is overexpressed in testicular seminoma (73). The role that overexpression of these initiation factors plays on the development and progression of cancer, if any, is not known.

**Overexpression of S6K.** S6K is amplified and highly overexpressed in the MCF7 breast cancer cell line, compared with normal mammary epithelium (74). In a study by Barlund *et al.* (74), S6K was amplified in 59 of 668 primary breast tumors, and a statistically significant association was observed between amplification and poor prognosis.

**Overexpression of PAP.** PAP catalyzes 3' poly(A) synthesis. PAP is overexpressed in human cancer cells compared with normal and virally transformed cells (75). PAP enzymatic activity in breast tumors has been correlated with PAP protein levels (76) and, in mammary tumor cytosols, was found to be an independent factor for predicting survival (76). Little is known, however, about how PAP expression or activity affects the translational profile.

**Alterations in RNA-binding Proteins.** Even less is known about alterations in RNA packaging in cancer. Increased expression and nuclear localization of the RNA-binding protein YB-1 are indicators of a poor prognosis for breast cancer (77), non-small cell lung cancer (78), and ovarian cancer (79). However, this effect may be mediated at least in part at the level of transcription, because YB-1 increases chemoresistance by enhancing the transcription of a multidrug resistance gene (80).

#### Activation of Signal Transduction Pathways

Activation of signal transduction pathways by loss of tumor suppressor genes or overexpression of certain tyrosine kinases can contribute to the growth and aggressiveness of tumors. An important mutant in human cancers is the tumor suppressor gene *PTEN*, which leads to the activation of the PI3K/Akt pathway. Activation of PI3K and Akt induces the oncogenic transformation of chicken embryo fibroblasts. The transformed cells show constitutive phosphorylation of S6K and of 4E-BP1 (81). A mutant Akt that retains kinase activity but does not phosphorylate S6K or 4E-BP1 does not transform fibroblasts, which suggests a correlation between the oncogenicity of PI3K and Akt and the phosphorylation of S6K and 4E-BP1 (81).

Several tyrosine kinases such as platelet-derived growth factor, insulin-like growth factor, HER2/neu, and epidermal growth factor receptor are overexpressed in cancer. Because these kinases activate downstream signal transduction pathways known to alter translation initiation, activation of translation is likely to contribute to the growth and aggressiveness of these tumors. Furthermore, the mRNA for many of these kinases themselves are under translational control. For example, HER2/neu mRNA is translationally controlled both by a short upstream open reading frame that represses HER2/neu translation in a cell type-independent manner and by a distinct cell type-dependent mechanism that increases translational efficiency (82). HER2/neu translation is different in transformed and normal cells. Thus, it is possible that alterations at the translational level can in part account for the discrepancy between *HER2/neu* gene amplification detected by fluorescence *in situ* hybridization and protein levels detected by immunohistochemical assays.

#### Translation Targets of Selected Cancer Therapy

Components of the translation machinery and signal pathways involved in the activation of translation initiation represent good targets for cancer therapy.

##### Targeting the mTOR Signaling Pathway: Rapamycin and Tumorstatin

Rapamycin inhibits the proliferation of lymphocytes. It was initially developed as an immunosuppressive drug for organ

transplantation. Rapamycin with FKBP 12 (FK506-binding protein, *M*, 12,000) binds to mTOR to inhibit its function.

Rapamycin causes a small but significant reduction in the initiation rate of protein synthesis (83). It blocks cell growth in part by blocking S6 phosphorylation and selectively suppressing the translation of 5' TOP mRNAs, such as ribosomal proteins, and elongation factors (83–85). Rapamycin also blocks 4E-BP1 phosphorylation and inhibits cap-dependent but not cap-independent translation (17, 86).

The rapamycin-sensitive signal transduction pathway, activated during malignant transformation and cancer progression, is now being studied as a target for cancer therapy (87). Prostate, breast, small cell lung, glioblastoma, melanoma, and T-cell leukemia are among the cancer lines most sensitive to the rapamycin analogue CCI-779 (Wyeth-Ayerst Research; Ref. 87). In rhabdomyosarcoma cell lines, rapamycin is either cytostatic or cytotoxic, depending on the p53 status of the cell; p53 wild-type cells treated with rapamycin arrest in the G<sub>1</sub> phase and maintain their viability, whereas p53 mutant cells accumulate in G<sub>1</sub> and undergo apoptosis (88, 89). In a recently reported study using human primitive neuroectodermal tumor and medulloblastoma models, rapamycin exhibited more cytotoxicity in combination with cisplatin and camptothecin than as a single agent. *In vivo*, CCI-779 delayed growth of xenografts by 160% after 1 week of therapy and 240% after 2 weeks. A single high-dose administration caused a 37% decrease in tumor volume. Growth inhibition *in vivo* was 1.3 times greater, with cisplatin in combination with CCI-779 than with cisplatin alone (90). Thus, preclinical studies suggest that rapamycin analogues are useful as single agents and in combination with chemotherapy.

Rapamycin analogues CCI-779 and RAD001 (Novartis, Basel, Switzerland) are now in clinical trials. Because of the known effect of rapamycin on lymphocyte proliferation, a potential problem with rapamycin analogues is immunosuppression. However, although prolonged immunosuppression can result from rapamycin and CCI-779 administered on continuous-dose schedules, the immunosuppressive effects of rapamycin analogues resolve in ~24 h after therapy (91). The principal toxicities of CCI-779 have included dermatological toxicity, myelosuppression, infection, mucositis, diarrhea, reversible elevations in liver function tests, hyperglycemia, hypokalemia, hypocalcemia, and depression (87, 92–94). Phase II trials of CCI-779 have been conducted in advanced renal cell carcinoma and in stage II/IV breast carcinoma patients who failed with prior chemotherapy. In the results reported in abstract form, although there were no complete responses, partial responses were documented in both renal cell carcinoma and in breast carcinoma (94, 95). Thus, CCI-779 has documented preliminary clinical activity in a previously treated, unselected patient population.

Active investigation is under way into patient selection for mTOR inhibitors. Several studies have found an enhanced efficacy of CCI-779 in PTEN-null tumors (30, 96). Another study found that six of eight breast cancer cell lines were responsive to CCI-779, although only two of these lines lacked PTEN (97). There was, however, a positive correlation between Akt activation and CCI-779 sensitivity (97). This correlation suggests that activation of the PI3K-Akt pathway,

regardless of whether it is attributable to a PTEN mutation or to overexpression of receptor tyrosine kinases, makes cancer cell amenable to mTOR-directed therapy. In contrast, lower levels of the target of mTOR, 4E-BP1, are associated with rapamycin resistance; thus, a lower 4E-BP1/eIF4E ratio may predict rapamycin resistance [98].

Another mode of activity for rapamycin and its analogues appears to be through inhibition of angiogenesis. This activity may be both through direct inhibition of endothelial cell proliferation as a result of mTOR inhibition in these cells or by inhibition of translation of such proangiogenic factors as vascular endothelial growth factor in tumor cells [99, 100].

The angiogenesis inhibitor tunicamycin, another anticancer drug currently under study, was also found recently to inhibit translation in endothelial cells [101]. Through a requisite interaction with integrin, tunicamycin inhibits activation of the PI3K/Akt pathway and mTOR in endothelial cells and prevents dissociation of eIF4E from 4E-BP1, thereby inhibiting cap-dependent translation. These findings suggest that endothelial cells are especially sensitive to therapies targeting the mTOR-signaling pathway.

#### Targeting eIF2 $\alpha$ EPA, Clotrimazole, mda-7, and Flavonoids

EPA is an n-3 polyunsaturated fatty acid found in the fish-based diets of populations having a low incidence of cancer [102]. EPA inhibits the proliferation of cancer cells [103], as well as in animal models [104, 105]. It blocks cell division by inhibiting translation initiation [105]. EPA releases Ca<sup>2+</sup> from intracellular stores while inhibiting their refilling, thereby activating PKR. PKR, in turn phosphorylates and inhibits eIF2 $\alpha$ , resulting in the inhibition of protein synthesis at the level of translation initiation. Similarly, clotrimazole, a potent antiproliferative agent *in vitro* and *in vivo*, inhibits cell growth through depletion of Ca<sup>2+</sup> stores, activation of PKR, and phosphorylation of eIF2 $\alpha$  [106]. Consequently, clotrimazole preferentially decreases the expression of cyclins A, E, and D1, resulting in blockage of the cell cycle in G<sub>1</sub>.

mda-7 is a novel tumor suppressor gene being developed as a gene therapy agent. Adenoviral transfer of mda-7 (Ad-mda7) induces apoptosis in many cancer cells including breast, colorectal, and lung cancer [107–109]. Ad-mda7 also induces and activates PKR, which leads to phosphorylation of eIF2 $\alpha$  and induction of apoptosis [110].

Flavonoids such as genistein and quercetin suppress tumor cell growth. All three mammalian eIF2 $\alpha$  kinases, PKR, heme-regulated inhibitor, and PERK/PEK, are activated by flavonoids, with phosphorylation of eIF2 $\alpha$  and inhibition of protein synthesis [111].

#### Targeting eIF4A and eIF4E: Antisense RNA and Peptides

Antisense expression of eIF4A decreases the proliferation rate of melanoma cells [112]. Sequestration of eIF4E by overexpression of 4E-BP1 is proapoptotic and decreases tumorigenicity [113, 114]. Reduction of eIF4E with antisense RNA decreases soft agar growth, increases tumor latency, and increases the rates of tumor doubling times [7]. Antisense eIF4E RNA treat-

ment also reduces the expression of angiogenic factors [115] and has been proposed as a potential adjuvant therapy for head and neck cancers, particularly when elevated eIF4E is found in surgical margins. Small molecule inhibitors that bind the eIF4G/4E-BP1-binding domain of eIF4E are proapoptotic [116] and are also being actively pursued.

#### Exploiting Selective Translation for Gene Therapy

A different therapeutic approach that takes advantage of the enhanced cap-dependent translation in cancer cells is the use of gene therapy vectors encoding suicide genes with highly structured 5' UTR. These mRNA would thus be at a competitive disadvantage in normal cells and not translate well, whereas in cancer cells, they would translate more efficiently. For example, the introduction of the 5' UTR of fibroblast growth factor-2 5' to the coding sequence of herpes simplex virus type-1 thymidine kinase gene, allows for selective translation of herpes simplex virus type-1 thymidine kinase gene in breast cancer cell lines compared with normal mammary cell lines and results in selective sensitivity to ganciclovir [117].

#### Toward the Future

Translation is a crucial process in every cell. However, several alterations in translational control occur in cancer. Cancer cells appear to need an aberrantly activated translational state for survival, thus allowing the targeting of translation initiation with surprisingly low toxicity. Components of the translational machinery, such as eIF4E, and signal transduction pathways involved in translation initiation, such as mTOR, represent promising targets for cancer therapy. Inhibitors of the mTOR have already shown some preliminary activity in clinical trials. It is possible that with the development of better predictive markers and better patient selection, response rates to single-agent therapy can be improved. Similar to other cytostatic agents, however, mTOR inhibitors are most likely to achieve clinical utility in combination therapy. In the interim, our increasing understanding of translation initiation and signal transduction pathways promise to lead to the identification of new therapeutic targets in the near future.

#### Acknowledgments

We thank Gayle Nesom from The University of Texas M. D. Anderson Cancer Center Department of Scientific Publications for editorial assistance and Dr. Elmer Bernstam for assistance with manuscript preparation.

#### References

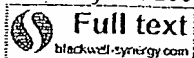
1. Pestova, T. V., Kolupaeva, V. G., Lomakin, I. B., Filipenko, E. V., Shatsky, I. N., Agol, V. I., and Hellen, C. U. Molecular mechanisms of translation initiation in eukaryotes. *Proc. Natl. Acad. Sci. USA*, 98: 7029–7036, 2001.
2. Rosenwald, I. B., Kaspar, R., Rousseau, D., Gehlke, L., Leboulch, P., Chen, J. J., Schmidt, E. V., Sonenberg, N., and London, I. M. Eukaryotic translation initiation factor 4E regulates expression of cyclin D1 at transcriptional and post-transcriptional levels. *J. Biol. Chem.*, 270: 21176–21180, 1995.
3. Rosenwald, I. B., Hutzler, M. J., Wang, S., Savas, L., and Fraire, A. E. Expression of eukaryotic translation initiation factors 4E and 2 $\alpha$  is increased frequently in bronchioloalveolar but not in squamous cell carcinomas of the lung. *Cancer (Phila.)*, 92: 2164–2171, 2001.



4. Darveau, A., Pelletier, J., and Sonenberg, N. Differential efficiencies of *in vitro* translation of mouse c-myc transcripts differing in the 5' untranslated region. *Proc. Natl. Acad. Sci. USA*, 82: 2315-2319, 1985.
5. Kozak, M. Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proc. Natl. Acad. Sci. USA*, 83: 2850-2854, 1986.
6. Kourmias, A. E., Lazaris-Karatzas, A., and Sonenberg, N. mRNAs containing extensive secondary structure in their 5' non-coding region translate efficiently in cells overexpressing initiation factor eIF-4E. *EMBO J.*, 11: 4153-4158, 1992.
7. Finkler-Schaeffer, C. W., Graff, J. R., De Benedetti, A., Zimmer, S. G., and Rhoads, R. E. Decreasing the level of translation initiation factor 4E with antisense RNA causes reversal of ras-mediated transformation and tumorigenesis of cloned rat embryo fibroblasts. *Int. J. Cancer*, 55: 841-847, 1993.
8. Kevil, C. G., De Benedetti, A., Payne, D. K., Coe, L. L., Laroux, F. S., and Alexander, J. S. Translational regulation of vascular permeability factor by eukaryotic initiation factor 4E: implications for tumor angiogenesis. *Int. J. Cancer*, 65: 785-790, 1996.
9. Nathan, C. A., Franklin, S., Abreo, F. W., Nassar, R., De Benedetti, A., and Glass, J. Analysis of surgical margins with the molecular marker eIF4E: a prognostic factor in patients with head and neck cancer. *J. Clin. Oncol.*, 17: 2909-2914, 1999.
10. Fukunaga, R., and Hunter, T. MNK1, a new MAP kinase-activated protein kinase, isolated by a novel expression screening method for identifying protein kinase substrates. *EMBO J.*, 16: 1921-1933, 1997.
11. Wasikiewicz, A. J., Flynn, A., Proud, C. G., and Cooper, J. A. Mitogen-activated protein kinases activate the serine/threonine kinases Mnk1 and Mnk2. *EMBO J.*, 16: 1909-1920, 1997.
12. Wang, X., Flynn, A., Wasikiewicz, A. J., Webb, B. L., Vries, R. G., Baines, I. A., Cooper, J. A., and Proud, C. G. The phosphorylation of eukaryotic initiation factor eIF4E in response to phorbol esters, cell stresses, and cytokines is mediated by distinct MAP kinase pathways. *J. Biol. Chem.*, 273: 9373-9377, 1998.
13. Pyronnet, S., Imataka, H., Gingras, A. C., Fukunaga, R., Hunter, T., and Sonenberg, N. Human eukaryotic translation initiation factor 4G (eIF4G) recruits Mnk1 to phosphorylate eIF4E. *EMBO J.*, 18: 270-279, 1999.
14. Klein, M., Schepfer, G. C., Voorma, H. O., and Thomas, A. A. Regulation of translation initiation factors by signal transduction. *Eur. J. Biochem.*, 253: 531-544, 1998.
15. Raught, B., and Gingras, A. C. eIF4E activity is regulated at multiple levels. *Int. J. Biochem. Cell Biol.*, 31: 43-57, 1999.
16. Takeuchi, K., Shibamoto, S., Nagamine, K., Stigemori, I., Omura, S., Kitamura, N., and Ito, F. Signaling pathways leading to transcription and translation cooperatively regulate the transient increase in expression of c-Fos protein. *J. Biol. Chem.*, 276: 26077-26083, 2001.
17. Kawasome, H., Papst, P., Webb, S., Keller, G. M., Johnson, G. L., Gelland, E. W., and Terada, N. Targeted disruption of p70(S6K) defines its role in protein synthesis and rapamycin sensitivity. *Proc. Natl. Acad. Sci. USA*, 95: 5033-5038, 1998.
18. Christie, G. R., Hajdich, E., Hundal, H. S., Proud, C. G., and Taylor, P. M. Intracellular sensing of amino acids in *Xenopus laevis* oocytes stimulates p70 S6 kinase in a target of rapamycin-dependent manner. *J. Biol. Chem.*, 277: 9952-9957, 2002.
19. Hara, K., Yonezawa, K., Weng, Q. P., Kozlowski, M. T., Betham, C., and Avruch, J. Amino acid sufficiency and mTOR regulate p70 S6 kinase and eIF-4E BP1 through a common effector mechanism. *J. Biol. Chem.*, 273: 14484-14494, 1998.
20. Graves, L. M., Bornfeldt, K. E., Argast, G. M., Krebs, E. G., Kong, X., Lin, T. A., and Lawrence, J. C., Jr. cAMP- and rapamycin-sensitive regulation of the association of eukaryotic initiation factor 4E and the translational regulator PHAS-I in aortic smooth muscle cells. *Proc. Natl. Acad. Sci. USA*, 92: 7222-7226, 1995.
21. Merrick, W. C., and Hershey, J. W. B. The pathway and mechanism of eukaryotic protein synthesis. In: J. W. B. Hershey and M. B. Mathews (eds.), *Translational Control*, pp. 31-69. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, 1996.
22. Kimball, S. R. Eukaryotic initiation factor eIF2. *Int. J. Biochem. Cell Biol.*, 31: 25-29, 1999.
23. Jagus, R., Joshi, B., and Barber, G. N. PKR, apoptosis and cancer. *Int. J. Biochem. Cell Biol.*, 31: 123-138, 1999.
24. Thomas, G., and Hall, M. N. TOR signaling and control of cell growth. *Curr. Opin. Cell Biol.*, 9: 782-787, 1997.
25. Gingras, A. C., Raught, B., and Sonenberg, N. Regulation of translation initiation by FRAP/mTOR. *Genes Dev.*, 15: 807-826, 2001.
26. Gingras, A. C., Gygi, S. P., Raught, B., Polakiewicz, R. D., Abraham, R. T., Hoekstra, M. F., Aebersold, R., and Sonenberg, N. Regulation of 4E-BP1 phosphorylation: a novel two-step mechanism. *Genes Dev.*, 13: 1422-1437, 1999.
27. Kumar, V., Pandey, P., Sabatini, D., Kumar, M., Majumder, P. K., Ehardt, A., Carmichael, G., Kufe, D., and Karbhandal, S. Functional interaction between RAFT1/FRAP/mTOR and protein kinase C $\delta$  in the regulation of cap-dependent initiation of translation. *EMBO J.*, 19: 1087-1097, 2000.
28. Yang, D. Q., and Kastan, M. B. Participation of ATM in insulin signaling through phosphorylation of eIF-4E-binding protein 1. *Nat. Cell Biol.*, 2: 893-898, 2000.
29. Liu, G., Zhang, Y., Bode, A. M., Ma, W. Y., and Dong, Z. Phosphorylation of 4E-BP1 is mediated by the p38/Msk1 pathway in response to UVB irradiation. *J. Biol. Chem.*, 277: 8810-8816, 2002.
30. Neshat, M. S., Mellingerhoff, I. K., Tran, C., Sikes, B., Thomas, G., Petersen, R., Frost, P., Gibbons, J. J., Wu, H., and Sawyers, C. L. Enhanced sensitivity of PTEN-deficient tumors to inhibition of FRAP/mTOR. *Proc. Natl. Acad. Sci. USA*, 98: 10314-10319, 2001.
31. Sekulic, A., Hudson, C. C., Homme, J. L., Yin, P., Ottensmeyer, D. M., Karnitz, L. M., and Abraham, R. T. A direct linkage between the phosphoinositide 3-kinase-AKT signaling pathway and the mammalian target of rapamycin in mitogen-stimulated and transformed cells. *Cancer Res.*, 60: 3504-3513, 2000.
32. Scott, P. H., and Lawrence, J. C., Jr. Attenuation of mammalian target of rapamycin activity by increased cAMP in 3T3-L1 adipocytes. *J. Biol. Chem.*, 273: 34496-34501, 1998.
33. Reynolds, I. T., Bodine, S. C., and Lawrence, J. C., Jr. Control of Ser2448 phosphorylation in the mammalian target of rapamycin by insulin and skeletal muscle load. *J. Biol. Chem.*, 277: 17657-17662, 2002.
34. Peterson, R. T., Beal, P. A., Comb, M. J., and Schreiber, S. L. FKBP12-rapamycin-associated protein (FRAP) autophosphorylates at serine 2481 under translationally repressive conditions. *J. Biol. Chem.*, 275: 7416-7423, 2000.
35. Peterson, R. T., Desai, B. N., Hardwick, J. S., and Schreiber, S. L. Protein phosphatase 2A interacts with the 70-kDa S6 kinase and is activated by inhibition of FKBP12-rapamycin-associated protein. *Proc. Natl. Acad. Sci. USA*, 96: 4438-4442, 1999.
36. McGrew, L. L., Dworkin-Rastl, E., Dworkin, M. B., and Richter, J. D. Poly(A) elongation during *Xenopus* oocyte maturation is required for translational recruitment and is mediated by a short sequence element. *Genes Dev.*, 3: 803-815, 1989.
37. Sheets, M. D., Wu, M., and Wickens, M. Polyadenylation of c-mos mRNA as a control point in *Xenopus* meiotic maturation. *Nature (Lond.)*, 374: 511-516, 1995.
38. Yamum, S. M., and Warrington, W. M. Deadenylation of maternal mRNAs during *Xenopus* oocyte maturation does not require specific cis-sequences: a default mechanism for translational control. *Genes Dev.*, 4: 2278-2286, 1990.
39. Gaffie, D. R. The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. *Genes Dev.*, 5: 2108-2116, 1991.
40. Sachs, A. B., and Varani, G. Eukaryotic translation initiation: there are (at least) two sides to every story. *Nat. Struct. Biol.*, 7: 356-361, 2000.
41. Woffe, A. P., and Meric, F. Coupling transcription to translation: a novel site for the regulation of eukaryotic gene expression. *Int. J. Biochem. Cell Biol.*, 28: 247-257, 1996.
42. Evdokimova, V. M., Wei, C. L., Sitkov, A. S., Simonenko, P. N., Lazarev, O. A., Vasilenko, K. S., Ustinov, V. A., Hershey, J. W., and Ovchinnikov, I. P. The major protein of messenger ribonucleoprotein particles in somatic cells is a member of the Y-box binding transcription factor family. *J. Biol. Chem.*, 270: 3186-3192, 1995.
43. Matsumoto, K., Meric, F., and Woffe, A. P. Translational repression dependent on the interaction of the *Xenopus* Y-box protein FRGY2 with mRNA: Role of the cold shock domain, tail domain, and selective RNA sequence recognition. *J. Biol. Chem.*, 271: 22706-22712, 1996.

44. Evdokimova, V., Ruzanov, P., Imataka, H., Raught, B., Svitkin, Y., Ovchinnikov, L. P., and Sonenberg, N. The major mRNA-associated protein YB-1 is a potent 5' cap-dependent mRNA stabilizer. *EMBO J.*, 20: 5491-5502, 2001.
45. Saito, H., Hayday, A. C., Wiman, K., Hayward, W. S., and Toneyawa, S. Activation of the c-myc gene by translocation: a model for translational control. *Proc. Natl. Acad. Sci. USA*, 80: 7476-7480, 1983.
46. Narbuz, C., Lafon, I., Audigier, S., Gensac, M. C., Vagner, S., Huez, G., and Prats, A. C. Alternative translation of the proto-oncogene c-myc by an internal ribosome entry site. *J. Biol. Chem.*, 272: 32061-32066, 1997.
47. Stoneley, M., Paulin, F. E., Le Quesne, J. P., Chappell, S. A., and Willis, A. E. c-Myc 5' untranslated region contains an internal ribosome entry segment. *Oncogene*, 16: 423-428, 1998.
48. Paulin, F. E., West, M. J., Sullivan, N. F., Whitney, R. L., Lyne, L., and Willis, A. E. Aberrant translational control of the c-myc gene in multiple myeloma. *Oncogene*, 13: 505-513, 1996.
49. Chappell, S. A., LeQuesne, J. P., Paulin, F. E., de Schootmeester, M. L., Stoneley, M., Soular, R. L., Ralston, S. H., Hellrich, M. H., and Willis, A. E. A mutation in the c-myc-IFES leads to enhanced internal ribosome entry in multiple myeloma: a novel mechanism of oncogene de-regulation. *Oncogene*, 19: 4437-4440, 2000.
50. Signori, E., Bagri, C., Papa, S., Primerano, B., Rinaldi, M., Amaldi, F., and Fazio, V. M. A. A somatic mutation in the 5'UTR of BRCA1 gene in sporadic breast cancer causes down-modulation of translation efficiency. *Oncogene*, 20: 4596-4600, 2001.
51. Liu, L., Diworith, D., Gao, L., Monzon, J., Summers, A., Lassam, N., and Hogg, D. Mutation of the CDKN2A 5' UTR creates an aberrant initiation codon and predisposes to melanoma. *Nat. Genet.*, 21: 128-132, 1999.
52. Savitsky, K., Platzer, M., Uziel, T., Gilad, S., Santiel, A., Rosenthal, A., Elroy-Steh, O., Shih, Y., and Rotman, G. Ataxia-telangiectasia structural diversity of untranslated sequences suggests complex post-transcriptional regulation of ATM gene expression. *Nucleic Acids Res.*, 25: 1678-1684, 1997.
53. Brown, C. Y., Mize, G. J., Pineda, M., George, D. L., and Morris, D. R. Role of two upstream open reading frames in the translational control of oncogene *mdm2*. *Oncogene*, 18: 5631-5637, 1999.
54. Sobczak, K., and Krzyzosiak, W. J. Structural determinants of BRCA1 translational regulation. *J. Biol. Chem.*, 277: 17349-17358, 2002.
55. Arick, B. A., Grendell, R. L., and Griffin, L. A. Enhanced translational efficiency of a novel transforming growth factor  $\beta$  mRNA in human breast cancer cells. *Mol. Cell. Biol.*, 14: 619-628, 1994.
56. Myers, S. M., Eng, C., Ponder, B. A., and Mulligan, L. M. Characterization of RET proto-oncogene 3' splicing variants and polyadenylation sites: a novel C-terminus for RET. *Oncogene*, 11: 2039-2045, 1995.
57. Byrne, J. A., Tomasello, C., Rouyer, N., Bellocq, J. P., Rio, M. C., and Bassel, P. The tissue inhibitor of metalloproteinases-3 gene in breast carcinoma: identification of multiple polyadenylation sites and a stromal pattern of expression. *Mol. Med.*, 1: 418-427, 1995.
58. Moscow, J. A., He, R., Gudas, J. M., and Cowan, K. H. Utilization of multiple polyadenylation signals in the human *RHOA* protooncogene. *Gene (Amst.)*, 144: 229-236, 1994.
59. Senterre-Lesentants, S., Alag, A. S., and Sobel, M. E. Multiple mRNA species are generated by alternate polyadenylation from the human *calmodulin-1* gene. *J. Cell. Biochem.*, 58: 445-454, 1995.
60. Lazaris-Karatzas, A., Montine, K. S., and Sonenberg, N. Malignant transformation by a eukaryotic initiation factor subunit that binds to mRNA 5' cap. *Nature (Lond.)*, 345: 544-547, 1990.
61. De Benedetti, A., and Rhoads, R. E. Overexpression of eukaryotic protein synthesis initiation factor 4E in HeLa cells results in aberrant growth and morphology. *Proc. Natl. Acad. Sci. USA*, 87: 8212-8216, 1990.
62. Polunovsky, V. A., Rosenwald, I. B., Tan, A. T., White, J., Chiang, L., Sonenberg, N., and Bitterman, P. B. Translational control of programmed cell death: eukaryotic translation initiation factor 4E blocks apoptosis in growth-factor-restricted fibroblasts with physiologically expressed or de-regulated Myc. *Mol. Cell. Biol.*, 16: 6573-6581, 1996.
63. Miyagi, Y., Sugiyama, A., Asai, A., Okazaki, T., Kuchino, Y., and Ken, S. J. Elevated levels of eukaryotic translation initiation factor eIF-4E mRNA in a broad spectrum of transformed cell lines. *Cancer Lett.*, 97: 247-252, 1995.
64. Kerekatte, V., Smiley, K., Hu, B., Smith, A., Gelder, F., and De Benedetti, A. The proto-oncogene/translation factor eIF4E: a survey of its expression in breast carcinomas. *Int. J. Cancer*, 64: 27-31, 1995.
65. Li, B. D., Liu, L., Dawson, M., and De Benedetti, A. Overexpression of eukaryotic initiation factor 4E (eIF4E) in breast carcinoma. *Cancer (Phila.)*, 79: 2385-2390, 1997.
66. Sorrells, D. L., Black, D. R., Meschonat, C., Rhoads, R., De Benedetti, A., Gao, M., Williams, B. J., and Li, B. D. Detection of eIF4E gene amplification in breast cancer by competitive PCR. *Ann. Surg. Oncol.*, 5: 232-237, 1998.
67. Li, B. D., McDonald, J. C., Nassar, R., and De Benedetti, A. Clinical outcome in stage I to III breast carcinoma and eIF4E overexpression. *Ann. Surg.*, 227: 756-761; discussion, 761-763, 1998.
68. Wang, S., Rosenwald, I. B., Hützel, M. J., Pihan, G. A., Savas, L., Chen, J. J., and Woda, B. A. Expression of the eukaryotic translation initiation factors 4E and 2a in non-Hodgkin's lymphomas. *Am. J. Pathol.*, 155: 247-255, 1999.
69. Fukuchi-Shimogori, T., Ishii, I., Kashiwagi, K., Mashiba, H., Ekimoto, H., and Igarashi, K. Malignant transformation by overproduction of translation initiation factor eIF4G. *Cancer Res.*, 57: 5041-5044, 1997.
70. Eberle, J., Krasagakis, K., and Orfanos, C. E. Translation initiation factor eIF-4A1 mRNA is consistently overexpressed in human melanoma cells in vitro. *Int. J. Cancer*, 71: 396-401, 1997.
71. Shuda, M., Kondoh, N., Tanaka, K., Ryo, A., Wakatsuki, T., Hada, A., Goseki, N., Igarashi, T., Hatsuse, K., Ahara, T., Honuchi, S., Shichita, M., Yamamoto, N., and Yamamoto, M. Enhanced expression of translation factor mRNAs in hepatocellular carcinoma. *Anticancer Res.*, 20: 2489-2494, 2000.
72. Nupponen, N. N., Pörkkä, K., Kakkola, L., Tanner, M., Persson, K., Borg, A., Isola, J., and Visakorpi, T. Amplification and overexpression of p40 subunit of eukaryotic translation initiation factor 3 in breast and prostate cancer. *Am. J. Pathol.*, 154: 1777-1783, 1999.
73. Rothe, M., Ko, Y., Albers, P., and Wernert, N. Eukaryotic initiation factor 3 p110 mRNA is overexpressed in testicular seminomas. *Am. J. Pathol.*, 157: 1597-1604, 2000.
74. Barklund, M., Forozan, F., Kononen, J., Bubendorf, L., Chen, Y., Bittner, M. L., Torhorst, J., Haas, P., Bucher, C., Sauter, G., Kallioniemi, O. P., and Kallioniemi, A. Detecting activation of ribosomal protein S6 kinase by complementary DNA and tissue microarray analysis. *J. Natl. Cancer Inst. (Bethesda)*, 92: 1252-1259, 2000.
75. Topfian, S. L., Kaneko, S., Gonzales, M. I., Bond, G. L., Ward, Y., and Manley, J. L. Identification and functional characterization of neo-poly(A) polymerase, an RNA processing enzyme overexpressed in human tumors. *Mol. Cell. Biol.*, 21: 5614-5623, 2001.
76. Scornilas, A., Talleri, M., Ardavanis, A., Courtis, N., Dimitriadis, E., Yotis, J., Tsiapalis, C. M., and Tringali, T. Polyadenylation polymerase enzymatic activity in mammary tumor cytosols: a new independent prognostic marker in primary breast cancer. *Cancer Res.*, 60: 5427-5433, 2000.
77. Jantz, M., Harbeck, N., Dettmar, P., Berger, U., Schmidt, A., Jurchott, K., Schmitt, M., and Royer, H. D. Y-box factor YB-1 predicts drug resistance and patient outcome in breast cancer independent of clinically relevant tumor biologic factors HER2, uPA and PAI-1. *Int. J. Cancer*, 97: 278-282, 2002.
78. Shibahara, K., Sugio, K., Osaki, T., Uchiyama, T., Maehara, Y., Kohno, K., Yasumoto, K., Sugimachi, K., and Kuwano, M. Nuclear expression of the Y-box binding protein, YB-1, as a novel marker of disease progression in non-small cell lung cancer. *Clin. Cancer Res.*, 7: 3151-3155, 2001.
79. Kamura, T., Yahata, H., Amada, S., Ogawa, S., Sonoda, T., Kobayashi, H., Mitsumoto, M., Kohno, K., Kuwano, M., and Nakano, H. Is nuclear expression of Y box-binding protein-1 a new prognostic factor in ovarian serous adenocarcinoma? *Cancer (Phila.)*, 85: 2450-2454, 1999.
80. Bargou, R. C., Jurchott, K., Wagener, C., Bergmann, S., Metzner, S., Bommer, K., Magera, M. Y., Winzer, K. J., Dietel, M., Dörken, B., and Royer, H. D. Nuclear localization and increased levels of transcription factor YB-1 in primary human breast cancers are associated with intrinsic MDR1 gene expression. *Nat. Med.*, 3: 447-450, 1997.
81. Aoki, M., Blazek, E., and Vogt, P. K. A role of the kinase mTOR in cellular transformation induced by the oncoproteins P3k and Akt. *Proc. Natl. Acad. Sci. USA*, 98: 136-141, 2001.
82. Child, S. J., Miller, M. K., and Geballe, A. P. Cell type-dependent and -independent control of HER-2/neu translation. *Int. J. Biochem. Cell Biol.*, 31: 201-213, 1999.

83. Jefferies, H. B., Reinhard, C., Kozma, S. C., and Thomas, G. Rapamycin selectively represses translation of the "polypyrimidine tract" mRNA family. *Proc. Natl. Acad. Sci. USA*, 91: 4441-4445, 1994.
84. Terada, N., Patel, H. R., Takase, K., Kohno, K., Naim, A. C., and Gelfand, E. W. Rapamycin selectively inhibits translation of mRNAs encoding elongation factors and ribosomal proteins. *Proc. Natl. Acad. Sci. USA*, 91: 11477-11481, 1994.
85. Jefferies, H. B., Fumagalli, S., Dennis, P. B., Reinhard, C., Pearson, R. B., and Thomas, G. Rapamycin suppresses 5'TOP mRNA translation through inhibition of p70s6k. *EMBO J.*, 16: 3693-3704, 1997.
86. Beretta, L., Gingras, A. C., Svitkin, Y. V., Hall, M. N., and Sonenberg, N. Rapamycin blocks the phosphorylation of 4E-BP1 and inhibits cap-dependent initiation of translation. *EMBO J.*, 15: 658-664, 1996.
87. Hidalgo, M., and Rowinsky, E. K. The rapamycin-sensitive signal transduction pathway as a target for cancer therapy. *Oncogene*, 19: 6680-6686, 2000.
88. Hosoi, H., Dilling, M. B., Shikata, T., Liu, L. N., Shiu, L., Ashman, R. A., Germain, G. S., Abraham, R. T., and Houghton, P. J. Rapamycin causes poorly reversible inhibition of mTOR and induces p53-independent apoptosis in human rhabdomyosarcoma cells. *Cancer Res.*, 59: 886-894, 1999.
89. Huang, S., and Houghton, P. J. Resistance to rapamycin: a novel anticancer drug. *Cancer Metastasis Rev.*, 20: 69-78, 2001.
90. Georger, B., Kerr, K., Tang, C. B., Fung, K. M., Powell, B., Sutton, L. N., Phillips, P. C., and Janss, A. J. Antitumor activity of the rapamycin analog CCI-779 in human primitive neuroectodermal tumor/medulloblastoma models as single agent and in combination chemotherapy. *Cancer Res.*, 61: 1527-1532, 2001.
91. Gibbons, J. J., Discalani, C., Peterson, R., Hernandez, R., Skotnicki, J., and Frost, P. The effect of CCI-779, a novel macrolide anti-tumor agent, on the growth of human tumor cells *in vitro* and in nude mouse xenografts *in vivo*. *Proc. Am. Assoc. Cancer Res.*, 40: 301, 1999.
92. Hidalgo, M., Rowinsky, E., Erlichman, C., Marshall, B., Marks, R., Edwards, T., and Buckner, J. J. A Phase I and pharmacological study of CCI-779 cycle inhibitor. *Ann. Oncol.*, 11 (Suppl. 4): 133, 2001.
93. Alexandre, J., Raymond, E., Depenbrock, H., Mekhaldi, S., Angevin, E., Paillet, C., Hanauske, A., Frisch, J., Feussner, A., and Amant, J. P. CCI-779, a new rapamycin analog, has antitumor activity at doses inducing only mild cutaneous effects and mucositis: early results of an ongoing Phase I study. *Proceedings of the 1999 AACR-NCI-EORTC International Conference*, Clin. Cancer Res., 5 (Suppl.): 3730s, 1999.
94. Chan, S., Johnston, S., Scheulen, M. E., Moss, K., Morant, A., Lahr, A., Feussner, A., Berger, M., and Kirsch, T. First report: a Phase 2 study of the safety and activity of CCI-779 for patients with locally advanced or metastatic breast cancer failing prior chemotherapy. *Proc. Am. Soc. Clin. Oncol.*, 21: 44a, 2002.
95. Atkins, M. B., Hidalgo, M., Stadler, W., Logan, T., Dutcher, J. P., Hudes, G., Park, Y., Marshall, B., Boril, J., and Dukart, G. A randomized double-blind Phase 2 study of intravenous CCI-779 administered weekly to patients with advanced renal cell carcinoma. *Proc. Am. Soc. Clin. Oncol.*, 21: 10a, 2002.
96. Smith, S. G., Trinh, C. M., Inge, L. J., Thomas, G., Cloughesy, T. F., Sawyers, C. L., and Mischel, P. S. PTEN expression status predicts glioblastoma cell sensitivity to CCI-779. *Proc. Am. Assoc. Cancer Res.*, 43: 335, 2002.
97. Yu, K., Toral-Barza, L., Discalani, C., Zhang, W. G., Skotnicki, J., Frost, P., and Gibbons, J. J. mTOR, a novel target in breast cancer: the effect of CCI-779, an mTOR inhibitor, in preclinical models of breast cancer. *Endocr. Relat. Cancer*, 8: 249-258, 2001.
98. Dilling, M. B., Germain, G. S., Dudkin, L., Jayaraman, A. L., Zhang, X., Harwood, F. C., and Houghton, P. J. 4E-binding proteins, the suppressors of eukaryotic initiation factor 4E, are downregulated in cells with acquired or intrinsic resistance to rapamycin. *J. Biol. Chem.*, 277: 13907-13917, 2002.
99. Guba, M., von Breitenbuch, P., Steinhilber, M., Koehl, G., Flegel, S., Hornung, M., Bruns, C. J., Zuelke, C., Farkas, S., Anthuber, M., Jauch, K. W., and Geissler, E. K. Rapamycin inhibits primary and metastatic tumor growth by antiangiogenesis: involvement of vascular endothelial growth factor. *Nat. Med.*, 8: 128-135, 2002.
100. Lane, H. A., Schell, C., Theuer, A., O'Reilly, T., and Wood, J. Anti-angiogenic activity of RAD001, an orally active anticancer agent. *Proc. Am. Assoc. Cancer Res.*, 43: 184, 2002.
101. Maeshima, Y., Suchak, A., Lively, J. C., Ueki, K., Kharbenda, S., Kahn, C. R., Sonenberg, N., Hynes, R. O., and Kalari, R. Tumstatin, an endothelial cell-specific inhibitor of protein synthesis. *Science (Wash. DC)*, 295: 140-143, 2002.
102. Caygill, C. P., Charlett, A., and Hill, M. J. Fat, fish oil and cancer. *Br. J. Cancer*, 74: 159-164, 1996.
103. Falconer, J. S., Ross, J. A., Fearon, K. C., Howkins, R. A., O'Flaherty, M. G., and Carter, D. C. Effect of eicosapentaenoic acid and other fatty acids on the growth *in vitro* of human pancreatic cancer cell lines. *Br. J. Cancer*, 69: 826-832, 1994.
104. Noguchi, M., Minami, M., Yagasaki, R., Kinoshita, K., Earashi, M., Kitagawa, H., Taniya, T., and Miyazaki, I. Chemoprevention of DMBA-induced mammary carcinogenesis in rats by low-dose EPA and DHA. *Br. J. Cancer*, 75: 348-353, 1997.
105. Palakurthi, S. S., Fluckiger, R., Aktas, H., Changolkar, A. K., Shahsafi, A., Hameit, S., Klic, E., and Halperin, J. A. Inhibition of translation initiation mediates the anticancer effect of the n-3 polyunsaturated fatty acid eicosapentaenoic acid. *Cancer Res.*, 60: 2919-2925, 2000.
106. Aktas, H., Fluckiger, R., Acosta, J. A., Savage, J. M., Palakurthi, S. S., and Halperin, J. A. Depletion of intracellular  $Ca^{2+}$  stores, phosphorylation of eIF2 $\alpha$ , and sustained inhibition of translation initiation mediate the anticancer effects of doxorubicin. *Proc. Natl. Acad. Sci. USA*, 95: 8280-8285, 1998.
107. Mhashikar, A. M., Schrock, R. D., Hindi, M., Liao, J., Sieger, K., Kourouma, F., Zou-Yang, X. H., Onishi, E., Takli, O., Vedrick, T. S., Fanger, G., Stewart, L., Watson, G. J., Snary, D., Fisher, P. B., Saeki, T., Roth, J. A., Ramesh, R., and Chada, S. Melanoma differentiation associated gene-7 (*mda-7*): a novel anti-tumor gene for cancer gene therapy. *Mol. Med.*, 7: 271-282, 2001.
108. Su, Z. Z., Madreddi, M. T., Lin, J. J., Young, C. S., Kilada, S., Reed, J. C., Goldstein, N. I., and Fisher, P. B. The cancer growth suppressor gene *mda-7* selectively induces apoptosis in human breast cancer cells and inhibits tumor growth in nude mice. *Proc. Natl. Acad. Sci. USA*, 95: 14400-14405, 1998.
109. Saeki, T., Mhashikar, A., Chada, S., Branch, C., Roth, J. A., and Ramesh, R. Tumor-suppressive effects by adenovirus-mediated *mda-7* gene transfer in non-small cell lung cancer cell *in vitro*. *Gene Ther.*, 7: 2051-2057, 2000.
110. Palaer, A., Vorburger, S. A., Barber, G. N., Chada, S., Mhashikar, A. M., Zou-Yang, H., Stewart, A. L., Balachandran, S., Roth, J. A., Hunt, K. K., and Swisher, S. G. Adenoviral transfer of the melanoma differentiation-associated gene 7 (*mda-7*) induces apoptosis of lung cancer cells via up-regulation of the double-stranded RNA-dependent protein kinase (PKR). *Cancer Res.*, 62: 2239-2243, 2002.
111. Ito, T., Wamken, S. P., and May, W. S. Protein synthesis inhibition by flavonoids: roles of eukaryotic initiation factor 2 $\alpha$  kinases. *Biochem. Biophys. Res. Commun.*, 265: 589-594, 1999.
112. Eberle, J., Fecker, L. F., Bittrich, J. U., Orfanos, C. E., and Geilen, C. C. Decreased proliferation of human melanoma cell lines caused by antisense RNA against translation factor eIF-4A1. *Br. J. Cancer*, 86: 1957-1962, 2002.
113. Polunovsky, V. A., Gingras, A. C., Sonenberg, N., Peterson, M., Tan, A., Rubins, J. B., Marivel, J. C., and Bitterman, P. B. Translational control of the antiapoptotic function of Ras. *J. Biol. Chem.*, 275: 24776-24780, 2000.
114. D'Cunha, J., Kratzke, M. G., Alter, M. D., Polunovsky, V. A., Bitterman, P. B., and Kratzke, R. A. Over-expression of the translational repressor 4E-BP1 inhibits NSCLC tumorigenicity *in vivo*. *Proc. Am. Assoc. Cancer Res.*, 43: 816-817, 2002.
115. DeFatta, R. J., Nathan, C. A., and DeBenedetti, A. Antisense RNA to eIF4E suppresses oncogenic properties of a head and neck squamous cell carcinoma cell line. *Laryngoscope*, 110: 928-933, 2000.
116. Herbert, T. P., Fahraeus, R., Prescott, A., Lane, D. P., and Proud, C. G. Rapid induction of apoptosis mediated by peptides that bind initiation factor eIF4E. *Curr. Biol.*, 10: 793-796, 2000.
117. DeFatta, R. J., Li, Y., and DeBenedetti, A. Selective killing of cancer cells based on translational control of a suicide gene. *Cancer Gene Ther.*, 9: 573-578, 2002.

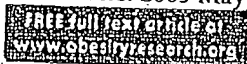


## Overexpression of chemokines, fibrogenic cytokines, and myofibroblasts in human membranous nephropathy.

Mezzano SA, Droguett MA, Burgos ME, Ardiles LG, Aros CA, Caorsi I, Egido J.

Division of Nephrology, School of Medicine, Universidad Austral, Valdivia, Chile.  
smezzano@uach.cl

Overexpression of chemokines, fibrogenic cytokines, and myofibroblasts in human membranous nephropathy. **BACKGROUND:** Proteinuria plays a central role in the progression of glomerular disease, and there is growing evidence suggesting that it may determine tubular cell activation with release of chemokines and fibrogenic factors, leading to interstitial inflammatory reaction. However, most studies on this subject have been performed in experimental models, and the experience in human kidney biopsies has been scarce. We analyzed the tissue sections of patients with idiopathic membranous nephropathy (IMN), a noninflammatory glomerular disease that may follow a progressive disease with heavy persistent proteinuria, interstitial cell infiltration, and decline of renal function. **METHODS:** Paraffin-embedded biopsy specimens from 25 patients with IMN (13 progressive and 12 nonprogressive) were retrospectively studied by immunohistochemistry [monocyte chemoattractant protein-1 (MCP-1), regulated on activation normal T-cell expressed and secreted chemokine (RANTES), osteopontin (OPN), platelet-derived growth factor-BB (PD-GF-BB)] and in situ hybridization [MCP-1, RANTES, PDGF-BB, transforming growth factor-beta1 (TGF-beta1)]. Moreover, we studied the presence of myofibroblasts, which were identified by the expression of alpha-smooth muscle actin (alpha-SMA), the monocytes/macrophages (CD68-positive cells), and T-cell infiltration (CD4+ and CD8+ cells). All of the patients were nephrotic and without treatment at time of the biopsy. **RESULTS:** A strong up-regulation of MCP-1, RANTES, and OPN expression was observed, mainly in tubular epithelial cells, with a significant major intensity in the progressive IMN patients. A strong correlation between the mRNA expression and the corresponding protein was noted. The presence of these chemokines and OPN was associated with interstitial cell infiltration. TGF-beta and PDGF were also up-regulated, mainly in tubular epithelial cells, with a stronger expression in the progressive IMN, and an association with the presence of myofibroblasts was found. **CONCLUSIONS:** Patients with severe proteinuria and progressive IMN have an overexpression in tubular epithelial cells of the chemokines MCP-1, RANTES, and OPN and the profibrogenic cytokines PDGF-BB and TGF-beta. Because this up-regulation was associated with an interstitial accumulation of mononuclear cells and an increase in myofibroblastic activity, it is suggested that those mediators are potential predictors of progression in IMN. Finally, based on experimental data and the findings of this article, we speculate that severe proteinuria is the main factor responsible for the up-regulation of these factors in tubular epithelial cells.



**Decreased uncoupling protein expression and intramyocytic triglyceride depletion in formerly obese subjects.**

Mingrone G, Rosa G, Greco AV, Manco M, Vega N, Hesselink MK, Castagneto M, Schräuwen P, Vidal H.

Istituto di Clinica Medica and. Clinica Chirurgica and Centro CNR Fisiopatologia Shock, Università Cattolica S Cuore, Rome, Italy. gmingrone@rm.unicatt.it

**OBJECTIVE:** To examine the muscular uncoupling protein expression 2 (UCP2) and UCP3 gene expression in morbid obese subjects before and after bariatric surgery [bilio-pancreatic diversion (BPD)]. **RESEARCH METHODS AND PROCEDURES:** Eleven obese subjects (BMI =  $49 \pm 2$  kg/m<sup>2</sup>) were studied before BPD and 24 months after BPD. Skeletal muscle UCP2 and UCP3 mRNA was measured using reverse transcriptase-competitive polymerase chain reaction and UCP3 protein by Western blotting. Intramyocytic triglycerides were quantified by high-performance liquid chromatography. Twenty-four-hour energy expenditure and respiratory quotient (RQ) were measured in a respiratory chamber. **RESULTS:** After BPD, the average weight loss was approximately 38%. Nonprotein RQ was increased in the postobese subjects ( $0.73 \pm 0.00$  vs.  $0.83 \pm 0.02$ ,  $p < 0.001$ ). The intramyocytic triglyceride level dropped ( $3.66 \pm 0.16$  to  $1.60 \pm 0.29$  mg/100 mg of fresh tissue,  $p < 0.0001$ ) after BPD. Expression of UCP2 and UCP3 mRNA was significantly reduced (from  $35.9 \pm 6.1\%$  to  $18.6 \pm 4.5\%$  of cyclophilin,  $p = 0.02$ ; from  $60.2 \pm 14.0\%$  to  $33.4 \pm 8.5\%$ ,  $p = 0.03$ ; respectively). UCP3 protein content was also significantly reduced ( $272.19 \pm 84.13$  vs.  $175.78 \pm 60.31$ , AU,  $p = 0.04$ ). A multiple regression analysis ( $R(2) = 0.90$ ) showed that IMTG levels ( $p = 0.007$ ) represented the most powerful independent variable for predicting UCP3 variation. **DISCUSSION:** The strong correlation of UCP expression and decrease in IMTG levels suggests that triglyceride content plays an even more important role in the regulation of UCP gene expression than the circulating levels of free fatty acids or the achieved degree of weight loss.

PMID: 12740453 [PubMed - indexed for MEDLINE]

**Differential expression of the short and long forms of the gamma 2 subunit of the GABAA/benzodiazepine receptors.**

Miralles CP, Gutierrez A, Khan ZU, Vitorica J, De Blas AL.

Division of Molecular Biology and Biochemistry, School of Biological Sciences,  
University of Missouri-Kansas City 64110-2499.

The distribution of the mRNAs encoding the gamma 2S and gamma 2L subunits of the GABAA receptor in the rat brain has been revealed by in situ hybridization, northern blot and dot blot analysis using specific antisense oligonucleotides. In addition, the quantitative distribution of the gamma 2S and gamma 2L subunit peptides participating in the fully assembled GABAA receptors/benzodiazepine receptors has been mapped by immunoprecipitation with specific anti-gamma 2S and anti-gamma 2L antibodies. Several neuronal types and brain regions are enriched in gamma 2L such as neurons of the layer II of striate cortex and cerebellar Purkinje cells as well as the inferior colliculus, superior colliculus, deep cerebellar nuclei, medulla and pons. Other neuronal types and regions are enriched in gamma 2S such as the mitral cells of the olfactory bulb, pyramidal neurons of the pyriform cortex, layer VI of the neocortex, granule cells of the dentate gyrus and pyramidal cells of the hippocampus. Other cortical areas and cerebellar granule cells express both gamma 2S and gamma 2L in comparable amounts. There is a good correlation between the relative expression of gamma 2S and gamma 2L mRNAs and the relative presence of these protein subunits in fully assembled and mature receptors in the studied brain regions. The differential distribution of gamma 2S and gamma 2L might result in differential ethanol sensitivity of the neurons expressing these GABAA receptor subunits.

PMID: 7968350 [PubMed - indexed for MEDLINE]

## Follicle-Stimulating Hormone Receptor and Its Messenger Ribonucleic Acid Are Present in the Bovine Cervix and Can Regulate Cervical Prostanoid Synthesis<sup>1</sup>

D. Mizrachi and M. Shemesh<sup>2</sup>

Department of Hormone Research, Kimron Veterinary Institute, Bet Dagan, Israel 50250

### ABSTRACT

The hypothesis that FSH regulates the bovine cervical prostaglandin E<sub>2</sub> (PGE<sub>2</sub>) synthesis that is known to be associated with cervical relaxation and opening at the time of estrus was investigated. Cervical tissue from pre-estrous/estrous, luteal, and postovulatory cows were examined for 1) the presence of bovine (b) FSH receptor (R) and its corresponding mRNA and 2) the effect of FSH on the PGE<sub>2</sub> regulatory pathway *in vitro*. The presence of bFSHR mRNA in the cervix (maximal during pre-estrous/estrus) was demonstrated by the expression of a reverse transcription (RT) polymerase chain reaction (PCR) product (384 base pairs) specific for bFSHR mRNA and sequencing. Northern blotting revealed three transcripts (2.5, 3.3, and 3.8 kilobases [kb]) in cervix from pre-estrous/estrous cows. The level of FSHR (75 kDa) was significantly higher ( $p < 0.01$ ) in Western blots of pre-estrous/estrous cervix than in other cervical tissues. There was a good correlation between the 75-kDa protein expression and its corresponding transcript of 2.55 kb throughout the estrous cycle as described by Northern blot analysis as well as RT-PCR. Incubation of FSH (10 ng/ml) with pre-estrous/estrous cervix resulted in a 3-fold increase in the expression of FSHR and a 2-fold increase in both G protein ( $\alpha$ ) and cyclooxygenase II. FSH (5–20 ng/ml) significantly increased ( $p < 0.01$ ) cAMP, inositol phosphate ( $p < 0.01$ ), and PGE<sub>2</sub> ( $p < 0.01$ ) production by pre-estrous/estrous cervix but not by cervix at the other stages. We conclude that bovine cervix at the time of the peripheral plasma FSH peak (pre-estrous/estrus) contains high levels of FSHR and responds to FSH by increasing the PGE<sub>2</sub> production responsible for cervical relaxation at estrus.

### INTRODUCTION

FSH and LH are the two known pituitary hormones that regulate gonadal functions. FSH is essential for female fertility, specifically for folliculogenesis, and human ovarian failure has been related to mutations both in the FSH receptor (FSHR) and in the FSH $\beta$  gene [1]. However, unlike LH, for which extra-gonadal binding has been documented for the uterus in the cow [2], pig [3], and human [4], no extra-gonadal effects of FSH have been reported. The uterine LH receptors have been shown to be physiologically active in the cow [2, 5]. The concentration of uterine LH receptors is related to the phase of the cycle, and binding of LH to uterine receptors increases uterine prostaglandin synthesis [2, 5]. Preliminary observations in our laboratory indicated that both FSH and LH receptors are also present in the cervix. We therefore investigated whether FSHRs are associated with the regulation of cervical prostaglandin, particularly around the time of estrus, when both cervical

PGE<sub>2</sub> production [6] and peripheral FSH concentrations are maximal.

The present investigation was therefore conducted to determine the presence of FSHR and its mRNA, and its physiological activity, in the bovine reproductive tract. Since the cervix is a prostaglandin E<sub>1</sub> (PGE<sub>1</sub>)- and PGE<sub>2</sub>-sensitive tissue and responds to hormonal stimulation, we also determined whether FSH could increase cervical cyclooxygenase expression and production of PGE<sub>2</sub> *in vitro*, particularly at the time near estrus.

### MATERIALS AND METHODS

#### Animals

Cervical and ovarian tissues were collected from Holstein cows at a local abattoir. The stage of the cycle was determined according to signs of ovulation and the status of corpora lutea (CL; weight, color). These parameters allowed the classification of the cervix into three groups: 1) pre-estrous/estrous (follicular, 18–20 days postovulation; regressed CL < 1.0 g; presence of cervical mucus;  $n = 54$ ); 2) postovulatory (1–4 days postovulation; presence of corpora hemorrhagica; signs of follicular rupture;  $n = 35$ ), and 3) luteal (12–16 days postovulation; CL 4–6 g;  $n = 44$ ). Bovine granulosa cells were collected from pre-estrous/estrous follicles and used as a positive control for FSHR expression [7].

The cervix external os segment was taken as the initial 3 cm of the cervix (total length of cervix is about 9 cm). The cervical os segments, primarily luminally oriented muscularis with epithelium, were sliced into horizontal strips (1 cm in length), pooled, and minced finely with a scalpel. The Kimron Veterinary Institute Animal Care and Use Committee approved all procedures.

#### FSHR Gene Expression

**RNA isolation.** Bovine cervical or ovarian tissues were rapidly dissected and frozen in liquid nitrogen (within 20 min after slaughter). Total RNA was extracted using the acid phenol, guanidine thiocyanate technique [8], using TriReagent (Molecular Research Center, Inc., Cincinnati, OH) according to the manufacturer's instructions.

**Reverse transcription (RT)-polymerase chain reaction (PCR).** Bovine FSHR mRNA was detected as previously described for the LH receptor mRNA [5]. Oligonucleotides corresponding to the published sequence of the bovine FSHR [7] were synthesized and used as primers in a PCR reaction. Bovine ovarian and cervical tissue cDNA obtained by RT of 2  $\mu$ g RNA was used as a template in the PCR reaction using a pair of primers corresponding to the transmembrane segments of the FSHR. The primers, 20-mer each, were selected using the OLIGO Program (Oligo R1 primer analysis software; National Biosciences, Plymouth, MN). The forward primer corresponded to position 1399–1419. (5'CGGCTTTTTCACCTGTCTTTG3') on the

Accepted April 26, 1999.

Received November 16, 1998.

<sup>1</sup>This work was supported by grant No. US-2333-93 from the US-Israel Binational Agricultural Research and Development Fund (BARD). This work is in partial fulfillment of a Ph.D. Thesis (D.M.) to be submitted to the Hadassah Medical School, Hebrew University.

<sup>2</sup>Correspondence: M. Shemesh, Department of Hormone Research, Kimron Veterinary Institute, Bet Dagan, POB 12, Israel 50250. FAX: 972 3 9681753; e-mail: mshem@vs.moag.gov.il

bovine FSHR mRNA [7]. The reverse primer was a 20-mer oligonucleotide complementary to position 1763–1783 (5'-CGCTTGGCTATCTTGGTGTC3'). The predicted size of the RT-PCR product was 384 basepairs (bp). Four microliters of the RT reaction were used as a template for PCR; the reaction was allowed to proceed 35 cycles using 2 U *Termus Aquaticus* DNA polymerase (recombinant) provided by MBI Fermentas, Vilnius, Lithuania, and 200 pmol of each primer. The cycling parameters of the PCR were 98°C for 20 sec, 56°C for 45 sec, and 72°C for 45 sec. After amplification the samples were separated on a 1% agarose gel, stained with ethidium bromide, and photographed under UV light. The 384-bp fragment was extracted using the Wizard PCR Preps Kit (Promega, Madison, WI) and sequenced with the upper primer using an automatic sequencer (Applied Biosystems, Foster City, CA). For internal control, a bovine  $\beta$ -actin fragment of 890 bp was produced from an upper primer (5'-ACCACTGGGACGACATGGAG3'; 21 mer) and a lower primer (5'-GCATTTGCGGTGGACAATGGA3'; 21 mer) as previously described [5]. Each PCR amplification was standardized using ovarian granulosa cDNA to produce a specific band of the predicted size. Standardization was based on temperature, cycle number, Mg level, and pH.

#### Western Blots

Cervical, bovine seminal vesicle, or muscle minces were sonicated for 45 sec, lysed in 500  $\mu$ l lysing buffer (1% Nonidet-40 [Sigma, St. Louis, MO], 2 mM EGTA, 5 mM  $MgCl_2$ , 1 mM PMSF in PBS) and incubated for 2 h at 4°C. After centrifugation, aliquots were taken for protein determination using gamma globulin (Sigma) as reference standard [9] and a protein dye binding method (Bio-Rad Laboratories, Richmond, CA). Sixty micrograms of protein of the soluble tissue extract was then separated by electrophoresis on 5–9% gradient SDS-PAGE and electroblotted onto nitrocellulose paper (Amersham, Little Chalfont, Bucks, UK) as described previously [9]. The nitrocellulose membrane was washed with PBS containing 0.05% TWEEN-20 and blocked with 10% horse serum (Kimron Veterinary Institute) in washing solution. The nitrocellulose membrane was then treated with either 1) an anti-peptide antibody (code name 179) raised against human (h) FSHR peptide 265–295, diluted 1:500 (donation of Dr. J.A. Dias, Wadsworth Center, New York State Department of Health); 2) rabbit antiserum for bovine G protein ( $\alpha_s$ ) diluted 1:1000; UBI, Lake Placid, NY); or 3) rabbit anti-bovine cyclooxygenase polyclonal antiserum (diluted 1:200; Kimron Veterinary Institute) [9]. Different dilutions were used for each antibody, and the final dilution used was 75% of the dilution, which gave a maximal signal as previously described [9]. The nitrocellulose paper was then incubated with horseradish peroxidase-conjugated goat anti-rabbit IgG (Sigma Israel, Rehovot, Israel; diluted 1:2000 in washing buffer) for 1 h at room temperature. The presence of FSHR, G protein ( $\alpha_s$ ), or cyclooxygenase was then visualized by means of a color reaction as follows. The nitrocellulose paper was incubated in a substrate solution containing 3,3'-diaminobenzidine (0.5 mg/ml; Sigma) in a mixture of PBS containing 0.5%  $CaCl_2$  and 6%  $H_2O_2$ . The antibody to hFSHR recognized the 75-kDa protein of the FSHR. The antibody to G ( $\alpha_s$ ) recognized both the 42-kDa and 87-kDa forms of this protein, and the antibody to cyclooxygenase recognized the 72-kDa form (cyclooxygenase II). Extract of bovine seminal vesicles that are known to have a high

content of cyclooxygenase was used as positive control as previously described [9]. The densitometric scans were obtained using a bio-imaging system (B.I.S. 2020; Rhenium Dingo, Jerusalem, Israel) and processed with Tina 2.0 software (Fuji, Japan). Linearity of detection [9] was determined for densitometry for both Western and Northern blots. Each Western blot was evaluated in the absence of the first antibody, and no signal was detected.

#### Northern Blot Analysis

For Northern blots, 10  $\mu$ g of total RNA was denatured in 20  $\mu$ l 50% formamide/2.2 M formaldehyde in single-strength 3-(*n*-morpholino) propanesulfonic acid (MOPS) buffer (0.04 M MOPS, 10 mM sodium acetate, 1 mM EDTA) at 65°C for 10 min. Samples were placed on ice, and 5  $\mu$ l of loading buffer (0.5% SDS, 0.25% bromophenol blue, 25% glycerol, 25 mM EDTA) was added. Samples were subjected to electrophoresis through a 1.2% agarose formaldehyde gel and transferred by capillary blotting to nylon membranes (Nytran; Schleicher and Schuell, Keene, NH). RNA was subsequently UV cross-linked to membranes. The RT-PCR product fragments of 384 bp for FSHR and 890 bp for  $\beta$ -actin were used to generate biotinylated probes using random primer biotin labeling of DNA for chemiluminescence (NEBlot phototype kit; Biolabs, Beverly, MA) according to the manufacturer's instructions. The blotted membrane was prehybridized at 65°C for 1 h in prehybridization solution (6-strength SSC [3 M NaCl, 0.3 M sodium citrate], 5-strength Denhardt's reagent [1% ficoll, 1% polyvinylpyrrolidone, 1% BSA], 0.5% SDS, and 100  $\mu$ g/ml denatured salmon sperm DNAs) and hybridized overnight in the same solution with the denatured biotinylated probe to the target RNA. The membrane was then washed in double-strength SSC, 0.1% SDS at room temperature for 10 min and washed again in 0.1-strength SSC, 0.1% SDS at 68°C for 15 min followed by subsequent chemiluminescence, development, and detection on x-ray film (RX; Fuji Film, Tokyo, Japan). To obtain quantitative data for the specific mRNA, we used the densitometric value for  $\beta$ -actin mRNA to normalize each specific mRNA value. To remove FSHR probe, membranes were stripped using water for 15 min, then incubated in 0.4 M NaOH, 0.1% SDS at 80°C for 30 min. The membrane was rinsed again in 0.2 M Tris-HCl, 0.1-strength SSC for 30 min at 25°C. After stripping, prehybridization and hybridization with  $\beta$ -actin probe were performed as described for FSHR mRNA.

#### Cyclic AMP Determinations

Cervical tissue minces (50 mg/ml) were incubated for 10 min in wells of 1.5 ml containing 1.0 ml of Tissue Culture Medium-199 (TCM-199 without serum (Biological Industries, Beit Haemek, Israel) in the absence or presence of physiological (5, 10 ng/ml) and pharmacological doses (20, 40 ng/ml) of FSH (USDA-bFSH-8-1; no further increase in cAMP production was seen at 50 or 100 ng/ml) or forskolin, a stimulator of adenylate cyclase (10  $\mu$ M; Sigma). The dose of forskolin was selected to give a 3-fold increase in cAMP. At the end of the incubation, tissues were removed, blotted on filter paper to remove mucus, and incubated overnight at 4°C with 400  $\mu$ l of 3%  $HClO_4$ . The solution was then neutralized with 150  $\mu$ l of  $KHCO_3$  and centrifuged, and 50- $\mu$ l aliquots were taken for the radioreceptor assay as described by Brown et al. [10] and modified for endometrium by Miyazaki et al. [11]. Standards (0–1000 pg of cAMP) or samples were incubated at room tem-



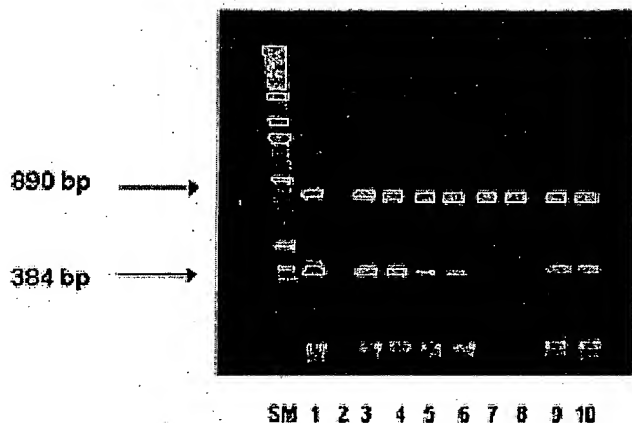


FIG. 1. RT-PCR amplification of bovine cervix FSHR and  $\beta$ -actin mRNA. RT-PCR was performed as described in *Materials and Methods* for FSHR and  $\beta$ -actin. Separate RT-PCR amplification products of FSHR (estimated 384 bp) and  $\beta$ -actin (estimated 890 bp) were combined and separated on agarose gel electrophoresis and stained with ethidium bromide. RT-PCR was performed using 2  $\mu$ g RNA of bovine granulosa cells as a positive control (lane 1), cervical tissue at the pre-estrous/estrous (lanes 3–4), post-ovulatory (lanes 5–6), and luteal phases (lanes 7–8). Lanes 9–10 represent reamplification, using as template the PCR product, cDNA, from the luteal-phase amplification. Lane 2 represents negative control in the absence of RNA. The tissue was taken from 4 cows at each stage of the estrous cycle, and two representative samples from each stage are shown. SM, size markers.

perature with 100  $\mu$ l of cAMP binding protein prepared from bovine adrenal extracts as described [10]. The buffer used for the assay was added to form a final reaction volume of 400  $\mu$ l. After 2-h incubation in a cold room, the reaction was stopped by adding 300  $\mu$ l of a charcoal-dextran solution; centrifugation followed, and the supernatant was removed for counting in a scintillation counter. The sensitivity of the assay was 60 pg/tube. Quadruplicate assays were made for each value determined. The within-assay and between-assay coefficients of variance were 8% and 10%, respectively.

1401  
 BTFSH GCTTTTTCAC TGTCTTTGCC AGTGAGCTCT CAGTCTACAC TCTGACCGCC  
 Cervix GCTTTTTCAC TGTCTTTACT AGTGAGCTCT CAGTCTACAC TCTGACCGCC  
  
 BTFSH ATCAGCGTGG AAAGATGGCA TACCATCACC CATGCCATGC AGCTCGAATG  
 Cervix ATCAGCGTGG AAAGATGGCA TCCATCACC CGTGCCATGC AGCTCGAATG  
  
 BTFSH CAAAGTGCGC CTCCGCCATG CTGCCAGCAT CATGCTGGTG GGCTGGATCT  
 Cervix CAAAGTGCGC CTCCGCCATG CTGCCAGCAT CATGCTGGTG GGCTGGATCT  
  
 BTFSH TTGCTTTTGC AGTTGCCCTT TTTCCCATCT TTGGCATCAG CAGCTACATG  
 Cervix TTGCTTTTGC AGTTGCCCTT TTTCCCATCT TTGGCATCAG CAGCTACATG  
  
 BTFSH AAGGTGAGCA TCTGCCTGCC CATGGACATT GACAGCCCCCT TGTCACAAC  
 Cervix AAGGTGAGCA TCTGCCTGCC CATGGACATT GACAGCCCCCT TGTCACAAC  
  
 BTFSH CTATGTCATG TCCCTCCTTG TGCTCAATGT CCTGGCCCTT GTGGTCATCT  
 Cervix CTATGTCATG TCCCTCCTTG TGCTCAATGT CCTGGCCCTT GTGGTCATCT  
  
 BTFSH GTGGGTGCTA CACTCACATC TACCTCACGG TGAGGAACCC CAACATCACA  
 Cervix GTGGGTGCTA CACTCACATC TACCTCACGG TGAGGAACCC CAACATCACA  
  
 1761  
 BTFSH TCCTCTCTA GTG  
 Cervix TCCTCTCTA GTG

FIG. 2. Automated nucleotide sequencing and homology between *Bos taurus* mRNA FSHR (GenBank accession number L22319) from nucleotides 1401 to 1763 and the PCR products obtained using the upper primer as described in *Materials and Methods*. A homology of 97.5% was found between *Bos taurus* and our amplified cDNA obtained by RT-PCR from cervical mRNA at pre-estrus/estrus, suggesting that they are complementary.

### Measurement of Phospholipase C (PLC) Activity

Cervical minces (30 mg) were incubated in 1 ml TCM-199 containing 5 mCi [ $^3$ H]myo-inositol for 90 min. Slices were rinsed with 1 ml TCM-199 and incubated for another 60 min to remove unincorporated [ $^3$ H]myo-inositol. At the end of the preincubations, slices were incubated in 1 ml TCM-199 containing 10 mM LiCl to inhibit inositol phosphate (IP) hydrolysis, allowing  $^3$ H-IP,  $^3$ H-IP<sub>2</sub>, and  $^3$ H-IP<sub>3</sub> to accumulate in the tissue after stimulation of PLC, in absence or presence of 0–20 ng/ml of FSH or 0–20 ng/ml LH (USDA bLH-I-1). Incubations were terminated by adding 2 ml ice-cold chloroform/methanol/hydrochloric acid (5:10:0.1, v:v:v) and then 1 ml chloroform and 1 ml EDTA (5 mM) and extracted as described by Kisielewska et al. [12].  $^3$ H-IP,  $^3$ H-IP<sub>2</sub>, and  $^3$ H-IP<sub>3</sub> were separated from the labeled inositol compounds by using Dowex AG 1-X8 (formate form).  $^3$ H-Labeled IP was eluted from the column by sequential elution into 10 fractions. Aliquots of each fraction were added to scintillation vials containing 5 ml scintillation fluid, and radioactivity was determined using a scintillation counter. Total activity of PLC is expressed as cpm of total IPs/30 mg/30 min.

### RIA for PGE<sub>2</sub> and PGF<sub>2 $\alpha$</sub>

Aliquots of 100  $\mu$ l were taken at the end of the incubation period for specific RIA of PGE<sub>2</sub> and PGF<sub>2 $\alpha$</sub> , which were performed without chromatographic separation. The antisera for PGF<sub>2 $\alpha$</sub>  (Sigma) reacts preferentially with PGF<sub>2 $\alpha$</sub>  but cross-reacts with PGF<sub>1 $\alpha$</sub>  (60%) and to a negligible extent (< 0.1%) with prostaglandins of the A, B, and E series. The antisera to PGE<sub>2</sub> (Sigma Israel) reacts preferentially with PGE<sub>2</sub> but cross-reacts with PGE<sub>1</sub> (20%), PGA<sub>1</sub>, PGA<sub>2</sub>, PGF<sub>1 $\alpha$</sub> , and PGF<sub>2 $\alpha$</sub>  (< 10%) and to a negligible extent with PGB<sub>1</sub> and PGB<sub>2</sub> (< 0.1%). The intraassay coefficients of variation were 9% and 11%, and the interassay coefficients of variance were 12.3% and 13% for PGF<sub>2 $\alpha$</sub>  and PGE<sub>2</sub>.

### Statistical Analysis

ANOVA was performed with a significance level of  $p < 0.05$ . Data were further analyzed using Tukey's procedure ( $p < 0.05$  or  $p < 0.01$ ) to assess significance between treatments. Student's  $t$ -test was used where appropriate. Values are expressed as means  $\pm$  SEM.

## RESULTS

### FSHR mRNA Expression in the Cervix

RT-PCR demonstrated that the FSHR gene is expressed in bovine cervix as indicated by the presence of the 384-bp fragment corresponding to bFSHR mRNA. The mRNA was maximally expressed in the cervix during the pre-estrous/estrous phase. In contrast, the expression of the gene was reduced in postovulation cervix and not detectable in luteal-phase cervix (Fig. 1). However, reamplification of the PCR product, cDNA, from the luteal phase produced a detectable signal (Fig. 1).

To demonstrate that the absence of FSHR mRNA in luteal-phase cervix was specific, bovine  $\beta$ -actin mRNA was used as internal control. PCR demonstrated that the specific 890 bp of the  $\beta$ -actin cDNA band was produced by the cervical tissues of all the stages. The band for  $\beta$ -actin was

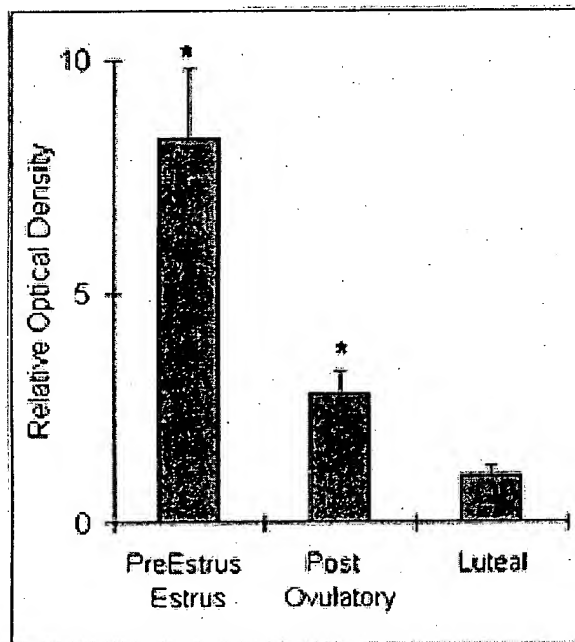
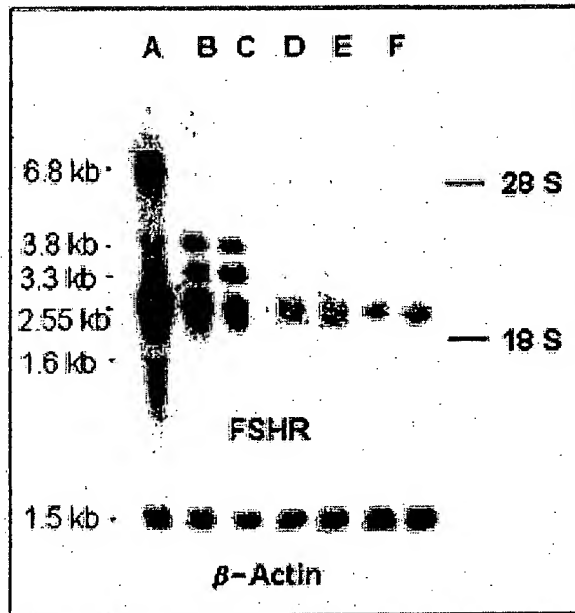


FIG. 3. Northern blot for analysis of total RNA (10  $\mu$ g) extracted from bovine cervical tissues collected from pre-estrus/estrus, postovulatory, and luteal-phase cows, and bovine granulosa cells as a positive control. Relative migration of 28S and 18S ribosomal RNA is indicated. To localize mRNA for FSHR and  $\beta$ -actin, our RT-PCR products (384-bp and 890-bp fragments, respectively) were used to generate biotinylated probes using biotin random primer. Figure shows results for FSHR mRNA in bovine granulosa cells as a positive control (lane A); and in cervical tissues at pre-estrus/estrus (lanes B-C), postovulation, (lanes D-E) and the luteal phase (lanes F-G). The corresponding  $\beta$ -actin band was obtained by re-probing of the membranes previously tested for FSHR mRNA as described in *Materials and Methods*. Columns show mean  $\pm$  SEM of the optical density of the ratio FSHR: $\beta$ -actin, in which the luteal phase is given an arbitrary value of 1.0. \*Significantly different from luteal-phase value ( $p < 0.01$ );  $n = 5$  for each stage.

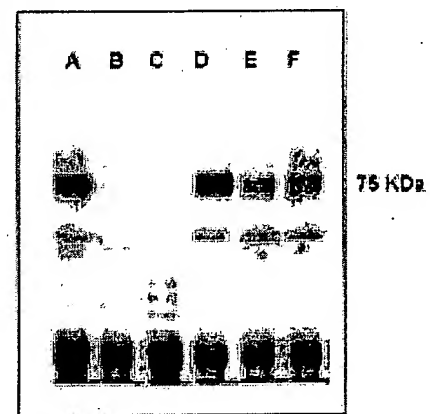
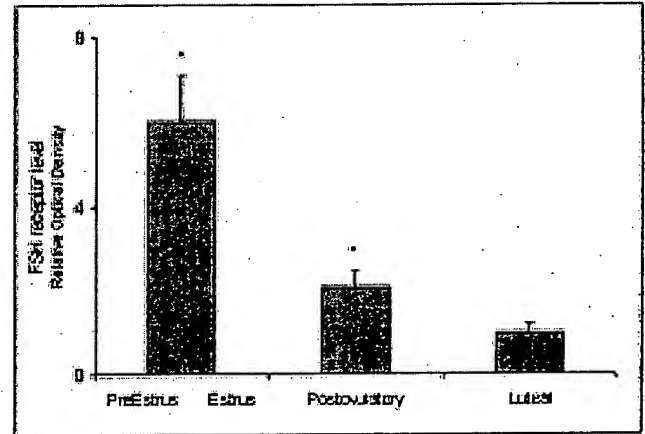


FIG. 4. Presence of FSHR protein in cervical tissues as determined by Western blot. Soluble cell extracts (60  $\mu$ g protein) were used to determine the amount of FSHR as described in *Materials and Methods*. Western blot proteins were transferred to nitrocellulose paper and probed with anti-serum to peptide hFSHR, 265-296. Western blot displays a representative result for bovine granulosa cells, used as positive control (lane A); bovine seminal vesicles (lane B) and muscle, used as a negative control (lane C); and cervical tissues from three stages of the estrous cycle: pre-estrus/estrus (lane D), postovulatory (lane E), and luteal (lane F). Histogram summarizes the results for the three stages of the estrous cycle. Columns show mean  $\pm$  SEM of the relative optical density in which the luteal phase is given an arbitrary value of 1.0. \*Significantly different from luteal-phase value ( $p < 0.01$ );  $n = 7$  for each stage.

present even though the 384-bp band corresponding to the FSHR was absent (Fig. 1).

The nucleotide sequence (Fig. 2) of the 384-bp fragment obtained from pre-estrus/estrus cervix was compared with that for the *Bos taurus* FSHR reported by Houde et al. [7]. It was found that there was a 97.5% homology between the 384-bp fragment and the nucleotide range 1410-1763 found in the *Bos taurus* FSHR.

The Northern blot contained the expected three transcripts of 2.55, 3.3, and 3.8 kilobases (kb) present in the bovine granulosa. However, the high (6.8-kb) and low (1.6-kb) transcripts found in the ovary were not found in the cervix. The three transcripts were found in pre-estrus/estrus cervix, but only one transcript (2.55 kb) was seen in luteal cervix (1/8 of the pre-estrus/estrus level;  $n = 5$ ) or postovulatory cervix (1/3 of the pre-estrus/estrus level;  $n = 5$ ) (Fig. 3).

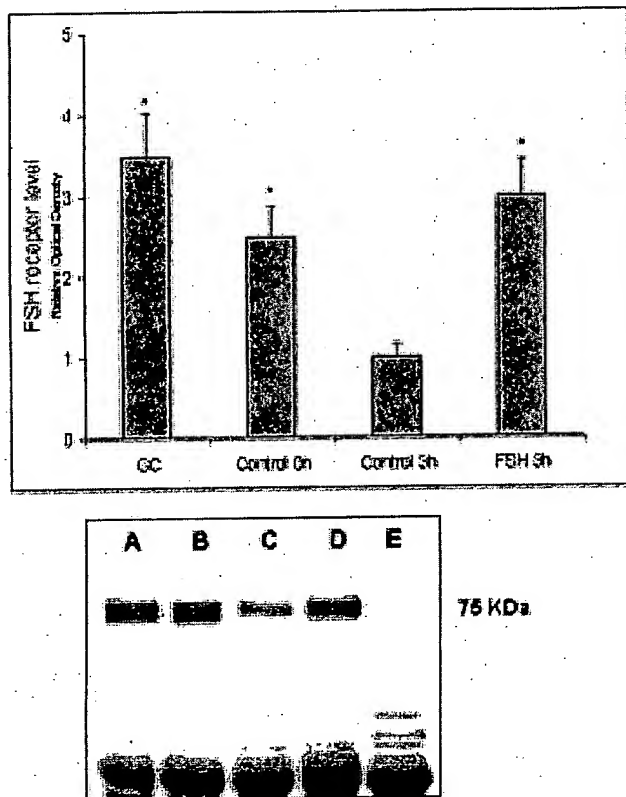


FIG. 5. Effect of incubation of cervical tissues with FSH on the expression of the FSHR protein. Cervical tissue from 5 pre-estrous/estrous cows was incubated in the absence or presence of FSH (0 or 10 ng/ml). Western blot is representative for FSH protein concentration after 3 h of incubation, with granulosa cell extract serving as positive control. Lane A, Granulosa cells (GC); lane B, cervical tissue at Time 0; lane C, cervical tissue after 3 h of incubation in TCM-199; lane D, cervical tissue after 3 h of incubation with 10 ng/ml of FSH. Lane E represents muscle tissue incubated with FSH (10 ng/ml) for 3 h as a negative control. Histogram summarizes the results of six experiments. Columns are means  $\pm$  SEM. \*Significantly different ( $p < 0.01$ ) from the column representing the control at 3 h (a given value of 1).

#### FSHR Protein in the Cervix

The presence of FSHR protein in the cervix was demonstrated as determined by Western blot analysis using a specific antipeptide antibody for hFSHR. The antibody recognized a 75-kDa protein in both bovine granulosa and cervical preparations identical with the predicted molecular size of FSHR. The signal for the 75-kDa protein was strongest in pre-estrous/estrous cervix when compared with postovulation- (3-fold) or luteal-phase (6-fold) cervix (Fig. 4). With the use of 60  $\mu$ g protein in each lane, this protein was observed to be expressed throughout the estrous cycle.

#### In Vitro Effect of FSH on Induction of FSHRs

A time-course (0.75, 1.5, and 3 h) and dose-response (0, 10, and 20 ng/ml) study was carried out to determine in vitro effect of FSH on its own receptor. Cervical minces (100 mg) from cows in the pre-estrous/estrous, postovulatory, and luteal phases were used. It was found, as determined by Western blot, that FSH induced its own receptor in cervical tissues at pre-estrous/estrus, with a 3-fold increase ( $p < 0.01$ ) observed at 10 ng/ml after 3 h of incubation (Fig. 5). However, no effect of FSH was seen at earlier

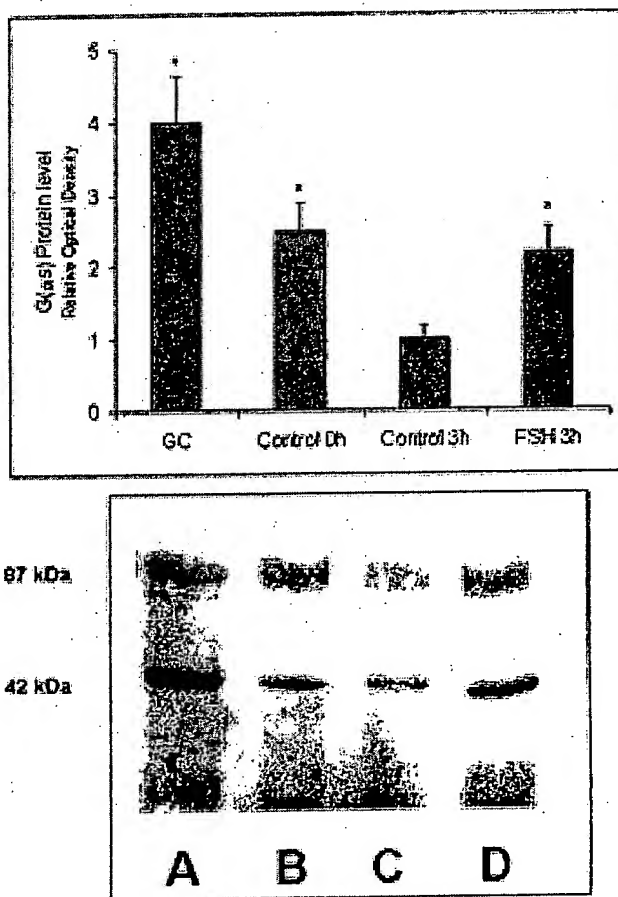


FIG. 6. Effect of incubation of cervical tissues with FSH on the synthesis of G protein. Cervical tissues from 5 pre-estrous/estrous cows were incubated in the absence or presence of FSH (10 ng/ml). Both the 87-kDa G protein ( $\alpha$ ,  $\beta$ ,  $\gamma$  complex) and 42-kDa protein ( $\alpha_s$ ) are elevated at 3 h of incubation. Granulosa cell extract was used as a positive control. Lane A, Granulosa cells (GC); lane B, cervical tissue at Time 0; lane C, cervical tissue after 3 h of incubation in TCM-199; lane D, cervical tissue after 3 h of incubation with 10 ng/ml of FSH. Columns are means  $\pm$  SEM of 42-kDa protein. \*Significantly different ( $p < 0.05$ ) from the column representing the control at 3 h (with an arbitrary value of 1.0).

times at the 10 ng/ml level. When 20 ng/ml was used, a stimulatory effect (2-fold) was seen after 1.5 h. No significant response was observed when tissues of the postovulatory or luteal phase were used at any dose or time tested (data not shown).

#### Effect of FSH Treatment on G Protein ( $\alpha_s$ )

A time-course and dose-response study was carried out to determine whether cervical FSHR was coupled to G protein. Cervical minces (100 mg) from cows in the pre-estrous/estrous, postovulatory, and luteal phases were incubated in the presence of FSH (0–20 ng/ml) for 0.75, 1.5, and 3 h. It was found in pre-estrous/estrous cervical tissue that, in the presence of FSH (10 ng/ml), there was a significantly ( $p < 0.05$ ) higher level of G protein ( $\alpha_s$ ) at the level of both the 87-kDa ( $\alpha$ ,  $\beta$ ,  $\gamma$  complex) and the 42-kDa ( $\alpha_s$ ) proteins after 3 h of incubation (Fig. 6). FSH, at either 10 or 20 ng/ml, did not have a stabilizing effect on the level of G protein ( $\alpha_s$ ) at shorter times of incubation (0.75 or 1.5 h). No significant response was observed when tis-

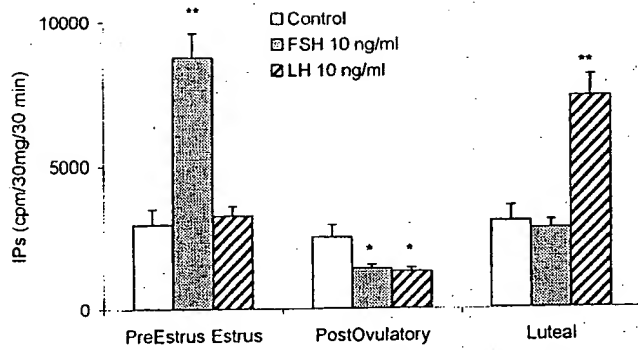


FIG. 7. Effect of FSH on IP production in cervical tissue. Cervical tissues from six cows at each of three stages of the cycle were incubated for 30 min in the absence or presence of FSH (10 ng/ml) or LH (10 ng/ml). IPs produced were determined in four replicates by amount of labeled compound (cpm) produced. Columns are means  $\pm$  SEM (\* $p$  < 0.05; \*\* $p$  < 0.01; significantly different from control value for each stage of the cycle).

sues of the postovulatory or luteal phase were used at any dose or time tested (data not shown).

#### IP Pathway Activation by FSH

For IP measurement, cervical tissue minces (30 mg) from pre-estrous/estrous, postovulatory, and luteal-phase cows were incubated for 30 min in the absence or presence of FSH (10 ng/ml) or LH (10 ng/ml). FSH significantly ( $p$  < 0.01) increased the level of IP, IP<sub>2</sub>, and IP<sub>3</sub> at pre-estrous/estrous. However, FSH significantly ( $p$  < 0.05) inhibited IP level in the postovulatory stage. In contrast, LH, but not FSH, stimulated IPs significantly ( $p$  < 0.01) at the luteal phase (Fig. 7).

#### Adenylate Cyclase Activation by FSH

A dose-response analysis was used to determine whether the cervical FSHR was coupled to adenylate cyclase. In these experiments, cervical tissue minces (50 mg) obtained from pre-estrous/estrous, postovulatory, and luteal-phase cows were incubated with FSH (0–40 ng/ml). FSH increased cAMP accumulation in a dose-dependent manner when incubated with pre-estrous/estrous cervical tissues, with maximal stimulation 2.5 times ( $p$  < 0.01) that of control in the presence of 10 ng/ml (Fig. 8A). In contrast, cervical tissue from luteal or postovulation did not respond to FSH even though tissues from all phases of the cycle responded to forskolin with a significant ( $p$  < 0.01) increase in cAMP (Fig. 8B).

#### Effect of FSH on PGE Production by Cervical Tissues

A dose-response analysis was carried out to determine whether FSHR was associated with cervical prostaglandin production. PGE<sub>2</sub> and PGF<sub>2 $\alpha$</sub>  production by cervical minces (100 mg/ml) in the presence of FSH (0–20 ng/ml) was determined by RIA. FSH stimulated ( $p$  < 0.01) basal PGE<sub>2</sub> production in a dose-dependent manner and caused a 3-fold increase in PGE<sub>2</sub> production at 10 ng/ml in pre-estrous/estrous cervical tissues ( $n$  = 8, Fig. 9). FSH had no effect on the small amount of basal PGF<sub>2 $\alpha$</sub>  (< 1 ng/100 mg) produced by cervical tissue (data not shown). FSH did not elevate PGE<sub>2</sub> in cervical tissues from the luteal phase; furthermore, a small but significant ( $p$  < 0.05) inhibition was observed at postovulation (Fig. 9).

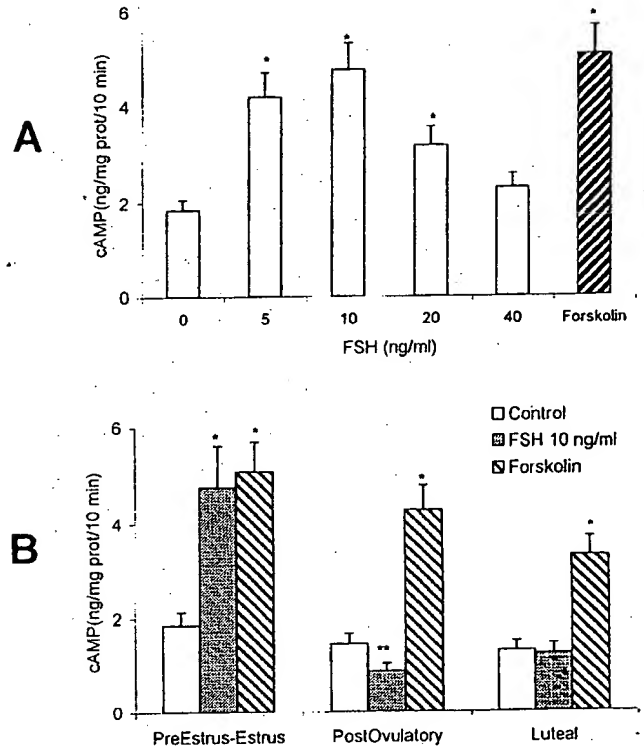


FIG. 8. Effect of FSH on adenylate cyclase production in cervical tissue. A) Cervical tissues from six cows at pre-estrous/estrous were incubated (4 replicates) for 10 min in the absence or presence of FSH (0, 5, 10, 20, or 40 ng/ml) or forskolin (10  $\mu$ mol/L), and the amount of cAMP produced was measured by a protein binding assay. B) Cervical tissues from six cows at each of the three stages of the estrous cycle were incubated (4 replicates) for 10 min in the absence or presence of FSH (10 ng/ml) or forskolin (10  $\mu$ M). The amount of cAMP produced was measured by a protein binding assay. Columns are means  $\pm$  SEM. Columns with asterisks were statistically different from their own control (\* $p$  < 0.01; \*\* $p$  < 0.05).

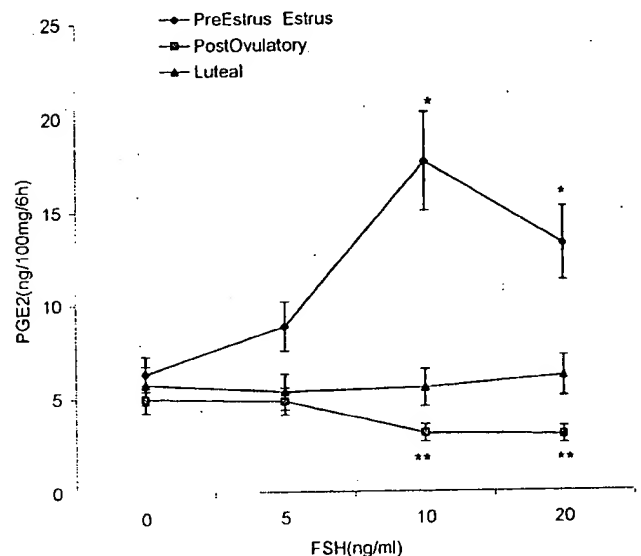


FIG. 9. FSH dose response on production of PGE<sub>2</sub> by cervical tissues. Cervical tissues from 8 pre-estrous/estrous, postovulatory, or luteal-phase cows were incubated in the absence or presence of FSH (5, 10, or 20 ng/ml) for 6 h, and the effect on the production of PGE<sub>2</sub> was measured by RIA. \* $p$  < 0.01; \*\* $p$  < 0.05.

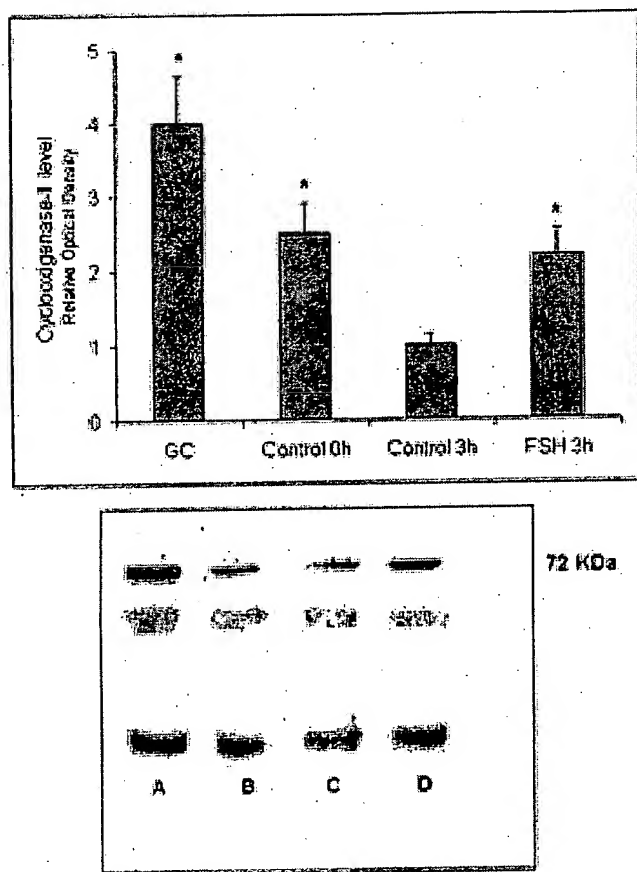


FIG. 10. Effect of incubation of cervical tissues with FSH on the expression of cyclooxygenase. Cervical tissues from 5 pre-estrous/estrous cows were incubated in the absence or presence of FSH (10 ng/ml). The 72-kDa cyclooxygenase II protein was elevated at 3 h of incubation. Lane A, bovine granulosa cells (GC); lane B, cervical tissue at Time 0; lane C, cervical tissue after 3 h of incubation in TCM-199; lane D, cervical tissue after 3 h of incubation with 10 ng/ml of FSH. Columns are means  $\pm$  SEM of 72-kDa protein. \*Significantly different ( $p < 0.01$ ) from the column representing the control at 3 h (with an arbitrary value of 1.0).

#### Cyclooxygenase Activation by FSH

To determine whether the presence of FSHR was associated with the expression of cyclooxygenase, cervical minces were incubated for 3 h in the presence or absence of FSH (10 ng/ml). Tissues from pre-estrous/estrous, post-ovulatory, and luteal-phase cows were extracted, separated on SDS-PAGE, and tested for the cyclooxygenase II (72 kDa) using a specific antibody. It was found that in cervical tissues from six pre-estrous/estrous cows, FSH induced a 200% increase in the expression of cyclooxygenase after 3 h of incubation (Fig. 10). In contrast, no effect on cyclooxygenase expression was seen in cervical tissues from luteal-phase cows, and an insignificant ( $p > 0.05$ ) elevation at the postovulatory phase was observed (data not shown).

#### DISCUSSION

Results of this study demonstrate by both PCR amplification and Northern blot analysis that FSHR mRNA is present in the bovine cervix. The bands corresponding to the receptor (75 kDa) and its mRNA were maximally expressed during the pre-estrous/estrous phase. Incubation of cervical tissues with FSH increased the expression of cyclooxygen-

ase and the production of  $PGE_2$  (and possibly  $PGE_1$ ) at the pre-estrous/estrous phase. The stimulation of cervical cyclooxygenase was associated with activation of both the PLC and adenylate cyclase second messenger G-protein related pathways.

Northern blot analysis of bovine cervical tissues from pre-estrous/estrous cows revealed multiple transcripts for bovine cervical FSHR. The sizes of the major transcripts were 2.5 kb, 3.3 kb, and 3.8 kb, which were similar to those obtained in bovine granulosa cells [7], rat testis and ovary [13, 14], and human ovary [15] and myometrium [16]. Other minor transcripts have been reported in all of these species. The RT-PCR product yielded a single band of 384 bp, which was the expected molecular size, so the different isoforms of the FSHR are probably the result of differential splicing of the same transcript. Isoforms of FSHR have been characterized in human [17] and primate [18] ovary and in ovine testis [19]. Since only a single 2.5-kb isomer was present in luteal-phase cervix (which did weakly produce FSHR), it would appear that there are different isoforms present in the bovine cervix as well.

The nucleotide sequence homology between our RT-PCR 384-bp product was 97.3% identical to the comparable region of the bovine FSHR, showing that our amplified cDNA was complementary to the *Bos taurus* mRNA FSHR. The PCR product of 384-bp cDNA was found primarily in the pre-estrous/estrous cervix.

Using specific antibody raised in rabbits against hFSHR (amino acid sequence 265–296) for Western blot analysis resulted in a major signal for a 75-kDa protein. A protein of similar molecular mass has been reported in rat and human ovary [20]. A strong signal for this protein was seen in the pre-estrous/estrous cervix compared with cervix from the other stages of the cycle. This cervical receptor was regulated by FSH itself, similar to the regulation of the FSHR in the rat ovary [14].

FSH increased cAMP production by cervical tissue from the pre-estrous/estrous phase but did not elevate cAMP in the postovulatory and luteal phases. FSH elevation of IPs displayed the same pattern. It would therefore appear that the cervical FSHR is associated with signal transduction pathways in a way similar to the accepted mechanism of LH action in the ovary and testes, i.e., activation of the second messenger pathways—adenylate cyclase [21] and phosphatidyl-inositol (PLC) [22, 23]. The inhibitory effect of FSH on IPs and cAMP at postovulation occur when FSHR is low. Since FSH activity is biphasic, i.e., is stimulated by a low dose and inhibited by a high dose, an inhibitory effect at low concentrations could be the result of the change in the ligand-receptor ratio.

FSH induced about a 200% increase in the expression of cyclooxygenase at pre-estrus/estrus but had no effect on the two other stages of the cycle. This was similar to the response of cyclooxygenase to LH that was observed for the bovine endometrium [2] and uterine vein [6]. However, in the endometrium, LH increased cyclooxygenase during both the luteal and postovulatory phases but not at estrus [2]. Furthermore, the induction of cyclooxygenase by gonadotropin in the endometrium was associated with an increased level of  $PGF_{2\alpha}$ , while in the cervix  $PGE_2$  was the major product.

The bFSH was of immunological grade and devoid of residual biological contamination of LH, thyroid-stimulating hormone, growth hormone, prolactin, or ACTH. The effects on increasing cAMP,  $PGE_2$ , and cyclooxygenase as well as induction of FSHR were therefore specific. Fur-

thermore, it was found that highly purified hFSH (10–20 ng/ml) (donation of J.A. Dias) had the same effects as bFSH when incubated with cervical tissue under the same conditions (2.5-fold increase in cAMP; 3-fold increase in PGE<sub>2</sub>; 2-fold increase in cyclooxygenase).

The relaxation and opening of the gravid cervix is due to active biochemical and structural changes in the cervical connective tissue that are mediated in part by prostaglandins [24]. It is thought that stretching of the cervix is a factor in causing the release of these prostaglandins [25, 26]. The effect of PGE in causing cervical softening in the ewe is also well documented [27–29]. In both nonpregnant and pregnant cows, PGE<sub>2</sub> causes an increase in cervical opening within 3 h of treatment [30]. PGE<sub>1</sub> administered intra-cervically in a jelly was shown to decrease cervical resistance within 24 h [31].

Recently, we reported [6] that oxytocin caused a significant stimulation of PGE<sub>2</sub> production in vitro in cervical tissues from pre-estrous/estrous cows but had no effect on PGE<sub>2</sub> production in cervical tissue from other stages of the estrous cycle of the cow. Similarly, in an initial report [31], it was shown that administration of oxytocin to pre-estrous cows increased the concentration of PGE<sub>2</sub> in the cervical exudate. However, peripheral oxytocin concentrations during estrus are lower than during the luteal phase [32, 33], and elevations of both peripheral oxytocin and cervical oxytocin receptor concentration are necessary for oxytocin to cause cervical softening towards parturition [24]. Furthermore, progesterone in vitro induced a dose-dependent inhibition of PGE<sub>2</sub> release by cervical tissues from pre-estrous/estrous cows, and this was associated with a decrease in both basal and oxytocin-stimulated PGE<sub>2</sub> production [6]. It would therefore seem that hormones other than oxytocin are responsible for the increase in cervical PGE<sub>2</sub> in the pre-estrous/estrous cow. The present report indicates that FSH, which has its peak peripheral concentration at the time of estrus, could be the hormone that increases cervical PGE<sub>2</sub>, as the FSHR expression is maximal at this time and FSH in vitro increases PGE<sub>2</sub> production by the cervix.

The bovine cervix at pre-estrous/estrous has high levels of FSHR protein and its corresponding mRNA. Activation of the receptor by FSH is associated with the G-protein-coupled receptor family that mediates the cAMP and IP signalling pathways. These signalling pathways then increase the expression of cyclooxygenase and production of PGE<sub>2</sub>. The expression of the FSHR was maximal at the time of the FSH peak in the blood and suggests a physiological role for FSH in the relaxation and opening of the cervix at estrus.

## ACKNOWLEDGMENTS

We thank Dr. J.A. Dias, Wadsworth Center, New York State Department of Health, for his donation of the anti-peptide antibody raised against hFSHR peptide 265–295 and purified hFSH; and the USDA Animal Hormone Program and NHPP for the bFSH and bLH.

## REFERENCES

- Dias JA, Lindau-Shepard B, Hauer C, Auger I. Human follicle-stimulating hormone structure-activity relationships. *Biol Reprod* 1998; 58:1331–1336.
- Freidman S, Gurevich M, Shemesh M. Bovine cyclic endometrium contains affinity luteinizing hormone/human chorionic gonadotropin. *Biol Reprod* 1995; 52:1020–1026.
- Ziecik AJ, Stanchev PD, Tilton JE. Evidence for the presence of luteinizing hormone/human chorionic gonadotropin-binding sites in the porcine uterus. *Endocrinology* 1986; 119:1159–1163.
- Reshef E, Lei ZM, Rao ChV, Pridham DD, Chegini N, Luborsky JL. The presence of gonadotropin receptors in nonpregnant human uterus, human placenta, fetal membranes, and decidua. *J Clin Endocrinol Metab* 1990; 70:421–430.
- Shemesh M, Gurevich M, Mizrahi D, Dombrovski L, Stram Y, Fields MJ, Shore LS. Expression of functional luteinizing hormone receptor and its mRNA in bovine uterine veins; luteinizing hormone induction of cyclooxygenase and augmentation of prostaglandin production in bovine uterine veins. *Endocrinology* 1997; 138:4844–4851.
- Shemesh M, Dombrovski L, Gurevich M, Friedman S, Shore LS, Fuchs AR, Fields MF. Regulation of bovine cervical secretion of prostaglandins and synthesis of cyclooxygenase by oxytocin. *Reprod Fertil Dev* 1997; 9:525–530.
- Houde A, Lambert A, Saumande J, Silversides DW, Lussier JG. Structure of bovine follicle-stimulating hormone receptor complementary DNA and expression in bovine tissues. *Mol Reprod Dev* 1994; 39:127–135.
- Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 1987; 162:156–159.
- Izhar M, Pasmanik M, Marcus S, Shemesh M. Dexamethasone inhibition of cyclooxygenase expression in bovine term placenta. *Prostaglandins* 1992; 43:239–254.
- Brown BL, Albano JDM, Ekins RP, Sgherzi AM, Tampion W. A simple and sensitive saturation assay method for measurement of adenosine 3',5'-cyclic monophosphate. *Biochem J* 1971; 121:561–562.
- Miyazaki K, Tanaka N, Kawakami S, Okamura H. Adenylyl cyclase activity in human decidua and myometrium during pregnancy and labour. *J Reprod Fertil* 1995; 105:141–145.
- Kisielewska J, Flint APF, Ziecik AJ. Phospholipase C and adenylyl cyclase signalling systems in the action of hCG on porcine myometrial smooth muscle cells. *J Endocrinol* 1996; 148:175–180.
- Heckert LL, Griswold MD. Expression of follicle-stimulating hormone receptor mRNA in rat testes and sertoli cells. *Mol Endocrinol* 1991; 5:670–677.
- LaPolt PS, Tilly JL, Aihara T, Nishimori K, Hsueh AJW. Gonadotropin-induced up-and-down-regulation of ovarian follicle-stimulating hormone (FSH) receptor gene expression in immature rats: effects of pregnant mare's gonadotropin, human chorionic gonadotropin, and recombinant FSH. *Endocrinology* 1992; 130:1289–1295.
- Tilly JL, Aihara T, Nishimori K, Jia X-C, Billig H, Kowalski KI, Perlas EA, Hsueh AJW. Expression of recombinant human follicle-stimulating hormone receptor: species-specific ligand binding, signal transduction, and identification of multiple ovarian messenger ribonucleic acid transcripts. *Endocrinology* 1992; 131:799–806.
- Korneye JL, Li X, Lei ZM, Rao ChV. Restoration of human chorionic gonadotropin response in human myometrial smooth muscle cells by treatment with follicle-stimulating hormone (FSH): evidence for the presence of FSH receptors in human myometrium. *Eur J Endocrinol* 1996; 134:225–231.
- Gromell J, Gutermann T, Nieschlag E. Molecular cloning of truncated isoform of the human follicle stimulating hormone receptor. *Biochem Biophys Res Commun* 1992; 188:1077–1083.
- Gromell J, Dankbar B, Sharma RS, Nieschlag E. Molecular cloning of the testicular follicle stimulating hormone receptor of the nonhuman primate *Macaca fascicularis* and identification of multiple transcripts in the testes. *Biochem Biophys Res Commun* 1993; 19:1066–1072.
- Khan H, Yarney TA, Sairam MR. Cloning of alternately spliced mRNA transcripts coding for variants of ovine testicular follitropin receptor lacking the G protein coupling domains. *Biochem Biophys Res Commun* 1993; 190:888–894.
- Liu X, DePasquale JA, Griswold MD, Dias JA. Accessibility of rat and human follitropin receptor primary sequence (R265-S296) in situ. *Endocrinology* 1994; 135:682–691.
- Marsh JM. The role of cyclic AMP in gonadal function. In: Greengard P, Robinson GA (eds.), *Cyclic Nucleotide Research*. Vol. 6. New York: Raven Press; 1975: 137–199.
- Davis JS. Stimulation of intracellular free Ca<sup>2+</sup> by luteinizing hormone in isolated bovine luteal cells. *Adv Exp Med Biol* 1987; 219: 671–675.
- Davis RJ. The mitogen-activated protein kinase signal transduction pathway. *J Biol Chem* 1993; 268:14553–14556.
- Fuchs A-R, Goeschen J, Rasmussen AB, Rehnström JV. Cervical ripening by endocervical and extra-amniotic PGE<sub>2</sub>. *Prostaglandins* 1984; 28:217–227.
- Ellwood DA, Mitchell MD, Anderson ABM, Turnbull AC. Specific changes in the in vitro production of prostanoids by the ovine cervix at parturition. *Prostaglandins* 1980; 39:675–684.

26. Hiller K, Coad N. Synthesis of prostaglandins by the human uterine cervix in vitro during passive mechanical stretch. *J Pharm Pharmacol* 1982; 34:262-263.
27. Stys SJ, Dresser BL, Ott TE, Clark KE. Effects of prostaglandin  $E_2$  on cervical compliance in pregnant ewes. *Am J Obstet Gynecol* 1981; 140:415-419.
28. Ledger WL, Ellwood DA, Taylor MJ. Cervical softening in late pregnant sheep by infusion of prostaglandin  $E_2$  into cervical artery. *J Reprod Fertil* 1983; 69:511-515.
29. Khalifa RME, Sayre BL, Lewis GS. Exogenous oxytocin dilates the cervix in the ewes. *J Anim Sci* 1992; 70:38-42.
30. Duchens M, Fredriksson G, Kindahl H, Aiumlamai S. Effect of intra-cervical administration of a prostaglandin  $E_2$  gel in pregnant and non-pregnant heifers. *Vet Rec* 1993; 133:546-549.
31. Graddy L, Fields M, Kowalski A, Chang S-M, Fuchs A-R. Effects of oxytocin and  $PGE_1$  on bovine cervix at estrus. *Biol Reprod* 1998; 58(suppl 1):174 (abstract 327).
32. Schams D. Oxytocin determination by radioimmunoassay. III. Improvement to subpicogram sensitivity and application to blood levels in cyclic cattle. *Acta Endocrinol* 1983; 103:180-183.
33. Ohtani M, Kobayashi S-I, Miyamoto A, Hayashi K, Yutaka F. Real-time relationships between intraluteal and plasma concentrations of endothelin, oxytocin and progesterone during prostaglandin  $F_{2\alpha}$ -induced luteolysis in the cow. *Biol Reprod* 1998; 58:103-108.

3: J Gen Virol. 2005 Oct;86(Pt 10):2769-80.

Related Articles, Links

Full text available at  
vir.sgmjournals.org

**The alpha(v)beta6 integrin receptor for Foot-and-mouth disease virus is expressed constitutively on the epithelial cells targeted in cattle.**

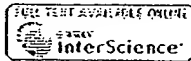
Monaghan P, Gold S, Simpson J, Zhang Z, Weinreb PH, Violette SM, Alexandersen S, Jackson T.

Institute for Animal Health, Pirbright Laboratory, Ash Road, Pirbright, Surrey GU24 0NF, UK.

Field strains of Foot-and-mouth disease virus (FMDV) use a number of alpha(v)-integrins as receptors to initiate infection on cultured cells, and integrins are believed to be the receptors used to target epithelial cells in animals. In this study, immunofluorescence confocal microscopy and real-time RT-PCR were used to investigate expression of two of the integrin receptors of FMDV, alpha(v)beta6 and alpha(v)beta3, within various epithelia targeted by this virus in cattle. These studies show that alpha(v)beta6 is expressed constitutively on the surfaces of epithelial cells at sites where infectious lesions occur during a natural infection, but not at sites where lesions are not normally formed. Expression of alpha(v)beta6 protein at these sites showed a good correlation with the relative abundance of beta6 mRNA. In contrast, alpha(v)beta3 protein was only detected at low levels on the vasculature and not on the epithelial cells of any of the tissues investigated. Together, these data suggest that in cattle, alpha(v)beta6, rather than alpha(v)beta3, serves as the major receptor that determines the tropism of FMDV for the epithelia normally targeted by this virus.

PMID: 16186231 [PubMed - in process]





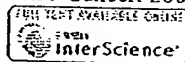
**Urokinase-mediated posttranscriptional regulation of urokinase-receptor expression in non small cell lung carcinoma.**

**Montuori N, Mattiello A, Mancini A, Taglialatela P, Caputi M, Rossi G, Ragno P.**

Istituto di Endocrinologia ed Oncologia Sperimentale, Consiglio Nazionale delle Ricerche, Naples, Italy.

The urokinase-type plasminogen activator (uPA) and its cellular receptor (uPAR) are involved in the proteolytic cascade required for tumor cell dissemination and metastasis, and are highly expressed in many human tumors. We have recently reported that uPA, independently of its enzymatic activity, is able to increase the expression of its own receptor in uPAR-transfected kidney cells at a posttranscriptional level. In fact, uPA, upon binding uPAR, modulates the activity and/or the level of a mRNA-stabilizing factor that binds the coding region of uPAR-mRNA. We now investigate the relevance of uPA-mediated posttranscriptional regulation of uPAR expression in non small cell lung carcinoma (NSCLC), in which the up-regulation of uPAR expression is a prognostic marker. We show that uPA is able to increase uPAR expression, both at protein and mRNA levels, in primary cell cultures obtained from tumor and adjacent normal lung tissues of patients affected by NSCLC, thus suggesting that the enzyme can exert its effect in lung cells. We investigated the relationship among the levels of uPA, uPAR and uPAR-mRNA binding protein(s) in NSCLC. Lung tissue analysis of 35 NSCLC patients shows an increase of both uPA and uPAR in tumor tissues, as compared to adjacent normal tissues, in 27 patients (77%); 19 of these 27 patients also show a parallel increase of the level and/or binding activity of a cellular protein capable of binding the coding region of uPAR-mRNA. Therefore, in tumor tissues, a strong correlation is observed among these 3 parameters; uPA, uPAR and the level and/or the activity of a uPAR-mRNA binding protein. We then suggest that uPA regulates uPAR expression in NSCLC at a posttranscriptional level by increasing uPAR-stability through a cellular factor that binds the coding region of uPAR-mRNA. Copyright 2003 Wiley-Liss, Inc.

PMID: 12704669 [PubMed - indexed for MEDLINE]



**Vascular endothelial growth factor expression correlates with matrix metalloproteinases MT1-MMP, MMP-2 and MMP-9 in human glioblastomas.**

**Munaut C, Noel A, Hougrand O, Foidart JM, Boniver J, Deprez M.**

Laboratory of Tumour and Development Biology, University of Liege, Liege, Belgium.

Vascular endothelial growth factor (VEGF) is the major endothelial mitogen in central nervous system neoplasms and it is expressed in 64-95% of glioblastomas (GBMs). Tumour cells are the main source of VEGF in GBMs whereas VEGF receptors (VEGFR-1, its soluble form sVEGFR-1, VEGFR-2 and neuropilin-1) are expressed predominantly by endothelial cells. Infiltrating tumour cells and newly-formed capillaries progress through the extracellular matrix by local proteolysis involving matrix metalloproteinases (MMPs). Recent studies have shown that VEGF expression and bioavailability can be modulated by MMPs. We reported previously that the expression of MT1-MMP in human breast cancer cells was associated with an enhanced VEGF expression. We used quantitative RT-PCR, Western blot, gelatin zymography and immunohistochemistry to study the expression of VEGF, VEGFR-1, VEGFR-2, sVEGFR-1, neuropilin-1, MT1-MMP, MMP-2, MMP-9 and TIMP-2 in 20 human GBMs and 5 normal brains. The expression of these MMPs was markedly increased in most GBMs with excellent correlation between mRNA and protein levels; activated forms of MMP-2 and MMP-9 were present in 8/18 and 7/18 of GBMs. A majority of GBMs (17/20) also expressed high levels of VEGF, as previously reported, with strong correlation between VEGF and MT1-MMP gene expression levels, and double-immunostaining showed that VEGF and MT1-MMP peptides co-localize in tumour and endothelial cells. Our results suggest that the interplay between metalloproteinases and VEGF previously described in experimental tumours may also be operative in human GBMs. Because of its dual ability to activate MMP-2 and to up-regulate VEGF, MT1-MMP might be of central importance in the growth of GBMs and represent an interesting target for anti-cancer treatments. Copyright 2003 Wiley-Liss, Inc.

PMID: 12918061 [PubMed - indexed for MEDLINE]



## VASCULAR ENDOTHELIAL GROWTH FACTOR EXPRESSION CORRELATES WITH MATRIX METALLOPROTEINASES MT1-MMP, MMP-2 AND MMP-9 IN HUMAN GLIOBLASTOMAS

Carine MUNAUT<sup>1</sup>, Agnès NOËL<sup>1</sup>, Olivier HOUGRAND<sup>2</sup>, Jean-Michel FOIDART<sup>1</sup>, Jacques BONIVER<sup>2</sup> and Manuel DEPREZ<sup>2\*</sup>

<sup>1</sup>Laboratory of Tumour and Development Biology, University of Liège, Liège, Belgium  
<sup>2</sup>Laboratory of Neuropathology, University of Liège, Liège, Belgium

Vascular endothelial growth factor (VEGF) is the major endothelial mitogen in central nervous system neoplasms and it is expressed in 64–95% of glioblastomas (GBMs). Tumour cells are the main source of VEGF in GBMs whereas VEGF receptors (VEGFR-1, its soluble form sVEGFR-1, VEGFR-2 and neuropilin-1) are expressed predominantly by endothelial cells. Infiltrating tumour cells and newly-formed capillaries progress through the extracellular matrix by local proteolysis involving matrix metalloproteinases (MMPs). Recent studies have shown that VEGF expression and bioavailability can be modulated by MMPs. We reported previously that the expression of MT1-MMP in human breast cancer cells was associated with an enhanced VEGF expression. We used quantitative RT-PCR, Western blot, gelatin zymography and immunohistochemistry to study the expression of VEGF, VEGFR-1, VEGFR-2, sVEGFR-1, neuropilin-1, MT1-MMP, MMP-2, MMP-9 and TIMP-2 in 20 human GBMs and 5 normal brains. The expression of these MMPs was markedly increased in most GBMs with excellent correlation between mRNA and protein levels; activated forms of MMP-2 and MMP-9 were present in 8/18 and 7/18 of GBMs. A majority of GBMs (17/20) also expressed high levels of VEGF, as previously reported, with strong correlation between VEGF and MT1-MMP gene expression levels, and double immunostaining showed that VEGF and MT1-MMP peptides co-localize in tumour and endothelial cells. Our results suggest that the interplay between metalloproteinases and VEGF previously described in experimental tumours may also be operative in human GBMs. Because of its dual ability to activate MMP-2 and to up-regulate VEGF, MT1-MMP might be of central importance in the growth of GBMs and represent an interesting target for anti-cancer treatments.  
© 2003 Wiley-Liss, Inc.

**Key words:** VEGF; MMPs; glioblastomas; brain; tumor

Angiogenesis is critical for the development of normal tissue and solid tumours. This process includes the degradation of the extracellular matrix (ECM) and the proliferation, migration and differentiation of endothelial cells, and is finely regulated by inhibitory and promoting factors.<sup>1,2</sup> Among positive factors, vascular endothelial growth factor (VEGF) has been proposed as the major endothelial mitogen in central nervous system (CNS) neoplasms.<sup>3</sup> Strong VEGF expression has been detected by immunohistochemistry in 64–95% of glioblastomas (GBMs).<sup>4–6</sup> GBMs are the most common malignant primitive tumours of the CNS in adults.<sup>7</sup> Microvascular proliferation is characteristic of these tumours and is an essential WHO diagnostic criteria.<sup>7</sup> Tumour cells are the main source of VEGF in GBMs whereas VEGF receptors are predominantly expressed by endothelial cells.<sup>8,9</sup> These receptors differ both in terms of affinity and transduction signaling.<sup>10–12</sup> VEGFR-1 (flt-1) and VEGFR-2 (KDR) belong to the Class II tyrosine-kinase receptor family. VEGFR-1 has a soluble isoform (sVEGFR-1) that modulates VEGF availability.<sup>8,13</sup> Neuropilin-1 (NRP1) is a co-receptor for VEGF that increases by 10-fold the affinity of the VEGFR<sub>2</sub> isoform for VEGFR-2.<sup>14</sup> NRP1 is thought to modulate VEGF-mediated tumour angiogenesis in human malignant astrocytomas.<sup>10</sup> The coordinated up-regulation of VEGF and its receptors appears as a critical event in the control of angiogenesis.<sup>8,12</sup>

Infiltrating tumour cells and newly-formed capillaries progress through the ECM by local proteolysis involving matrix metalloproteinases (MMPs).<sup>15,16</sup> MMPs are proteolytic enzymes that are synthesized as inactive zymogens. Their activation requires the removal of a propeptide by proteinase cleavage and can be inhibited by various tissue inhibitors of MMPs (TIMPs). Most MMPs are secreted as soluble enzymes but a subset of them are inserted in the cell membrane by a transmembrane domain or by a glycosylphosphatidyl-inositol anchor and are classified as membrane-type MMPs (MT-MMPs).<sup>17,18</sup> In cultured tumour cells, MT-MMPs tend to accumulate on the cytoplasmic membrane of invadopodia where they selectively mediate local pericellular proteolysis. GBMs express high levels of MT1-MMP, MMP-2 and MMP-9.<sup>19,20</sup> Among these MMPs, MT1-MMP might play a central role in the remodeling of the ECM as this membrane-bound protease is able to activate MMP-2 and MMP-13.<sup>21,22</sup> Moreover, MT1-MMP has been shown to promote cell migration in various carcinoma cell lines by its ability to cleave laminin-5, a major constituent of basement membrane,<sup>23,24</sup> and through the processing of CD44H (the major receptor for hyaluronan)<sup>25</sup> and of  $\alpha_5\beta_3$  integrin.<sup>26</sup>

MT1-MMP is involved in both developmental and tumour angiogenesis.<sup>27</sup> MT1-MMP overexpression in human melanoma cells has been associated with enhanced *in vitro* invasion and increased *in vivo* tumour growth and vascularization.<sup>28</sup> We have shown previously that in MCF-7 breast cancer cells, VEGF transcription is upregulated when MT1-MMP is overexpressed.<sup>29</sup> Up-regulation of VEGF by MT1-MMP has also been reported in a model of human glioma xenograft by Deryugina *et al.*<sup>30</sup> These experimental data suggest a link between MT1-MMP and the VEGF network. We have tested for the presence of such a link in human glioblastomas. We compared the expression of VEGF and its receptors (VEGFR-1, sVEGFR-1, VEGFR-2, NRP1) with

Grant sponsor: Communauté Française de Belgique (Actions de Recherches Concertées); Grant sponsor: Commission of European Communities; Grant sponsor: Fonds de la Recherche Scientifique Médicale; Grant sponsor: Fonds National de la Recherche Scientifique (FNRS, Belgium); Grant sponsor: Fédération Belge Contre le Cancer; Grant sponsor: Fonds Spéciaux de la Recherche (University of Liège); Grant sponsor: Centre Anticancéreux près l'Université de Liège; Grant sponsor: FB Assurances; Grant sponsor: Fondation Léon Frédéricq (University of Liège); Grant sponsor: D.G.T.R.E. from the "Région Wallonne"; Grant sponsor: Fonds d'Investissements de la Recherche Scientifique (CHU, Liège, Belgium); Grant sponsor: Federal Office for Scientific, Technical and Cultural Affairs (Brussels, Belgium).

\*Correspondence to: Laboratory of Neuropathology, Tower of Pathology B23/1, B-4000 Sart Tilman (Liège), Belgium. Fax: +32-4-3662919. E-mail: Manuel.Deprez@ulg.ac.be

Received 23 October 2002; Revised 12 March 2003; Accepted 27 April 2003

DOI 10.1002/ijc.11313  
Published online 26 June 2003 in Wiley InterScience (www.interscience.wiley.com).

This material may be protected by Copyright law (Title 17 U.S. Code)

MT1-MMP, MMP-2, MMP-9 and TIMP-2 in a series of 20 GBMs and 5 normal brains. Using quantitative RT-PCR, gelatin zymography, Western blot and immunohistochemistry, we showed a strong correlation between the expression of VEGF, MT1-MMP, MMP-2 and MMP-9 in GBMs. These results are in accordance with previous *in vitro* studies and add to the evidence of an interplay between VEGF and MMPs in the progression of human GBMs.

#### MATERIAL AND METHODS

##### Patients

We studied 20 GBMs diagnosed at the Laboratory of Neuro-pathology-CHU Liège between 1997 and 2001. The series included 17 primary GBMs (i.e., no previous history of lower grade diffuse astrocytoma) and 3 secondary GBMs (i.e., previous history of lower grade diffuse astrocytoma). Clinical information on these 20 cases have been reported previously as part of a larger series.<sup>31</sup> The gender ratio was 1/1, and the age at time of diagnosis ranged from 41–79 years (mean 56 years). Normal brain cortex and white matter were obtained from 5 patients with intractable epilepsy treated by partial temporal lobectomy. Histological examination of these specimens showed severe hippocampal sclerosis; frozen tissue was sampled from microscopically normal inferior temporal gyri. Our study was approved by the Ethical Committee of the Faculty of Medicine of the University of Liège.

##### RNA extraction and cDNA synthesis

Total RNA was extracted from cryosections with RNeasy Mini Kit (QIAGEN GmbH, Hilden, Germany) according to the manufacturer's protocol. Total RNA (1 µg) was reverse transcribed with a ThermoScript reverse transcriptase (ThermoScript RT-PCR System, Invitrogen, Carlsbad, CA) and random hexamers as primers.

##### Primers

Primers pairs used in our study are described in Table I. Primers for the VEGF gene were chosen to distinguish between VEGF<sub>189</sub>, VEGF<sub>165</sub>, VEGF<sub>145</sub> and VEGF<sub>121</sub> mRNA isoforms. Intron-spanning primers and probes for the TaqMan system (primers for VEGFR-1 (Flt-1), sVEGFR-1, VEGFR-2 (KDR/Flk-1) and NRP1) were designed to meet specific criteria by using Primer Express software (Perkin Elmer, Foster City, CA). All primers were synthesized by Eurogentec (Liège, Belgium). The 5'- and 3'-end nucleotides of the probe were labeled with a reporter (FAM = 6-carboxy-fluorescein) and a quencher dye (TAMRA = 6-carboxy-tetramethylrhodamine). We conducted BLASTn (National Center for Biotechnology Information, Bethesda) searches against dbEST and the non redundant set of GenBank, EMBL, and DDBJ database sequences to confirm the total gene specificity of the nucleotide sequences chosen for the primers. The specificity of the amplified PCR products was confirmed either by restriction digest or by sequencing. The 18S ribosomal RNA was measured using the Pre-Developed TaqMan Assay Reagents Endogenous control kit from Applied Biosystems (Foster City, CA).

TABLE I—SEQUENCE OF PRIMERS AND TaqMan PROBES USED FOR RT-PCR STUDIES

Gene and accession number	Position	Sequence	Size	Cycles
MMP-2 FP	1740F	5'-AGATCTTCTCTCTCAAGGACCGGT-3'	225 bp	33
MMP-2-RP NM_004530	1964R	5'-GGCTGGTCAGTGGCTTGGGGTA-3'		
MMP-9-FP	1592F	5'-GCGGAGATTGGGAACAGCTGTA-3'	208 bp	37
MMP-9-RP J05070	1800R	5'-GACGCGCCTGTGTACACCCACA-3'		
MMP-14-FP	1288F	5'-GGATACCCAATGCCCATTTGGCCA-3'	221 bp	32
MMP-14-RP NM_004995	1508R	5'-CCATTGGGCATCCAGAAGAGAGAC-3'		
TIMP-1-FP	78F	5'-CATCCTGTGTGTGCTGTGGCTGAT-3'	168 bp	33
TIMP-1-RP M12670	245R	5'-GTCATCTTGATCTCATAACGCTGG-3'		
TIMP-2-FP	78F	5'-CTCGCTGGACGTTGGAGGAAAGAA-3'	155 bp	30
TIMP-2-RP NM_003255	245R	5'-AGCCCATCTGGTACCTGTGGTTCA-3'		
VEGF-FP	1208F	5'-CCTGGTGGACATCTTCCAGGAGTA-3'	479 bp 407 bp 347 bp 275 bp	33
VEGF-RP AH001553	1687R	5'-CTCACC GGCTCGGCTTGTCACA-3'		
28S rRNA-RP	12403F	5'-GTTCAACCCACTAATAGGGAACGTGA-3'		
28S rRNA-RP U13369	12614R	5'-GATCTGACTTAGAGGCGTTCACT-3'		
VEGFR1-FP	2438F	5'-TCCCTTATGATGCCAGCAAGT-3'	79 bp	40
VEGFR1-RP VEGFR1 Probe AF063657	2516R 2469	5'-CCAAAAGCCCCCTCTTCCAA-3' 5'-CCGGGAGAGACTTAAACTGGCAATCA-3'		
sVEGFR1-FP	2209F	5'-ACAATCAGAGGTGAGCACTGCAA-3'	180 bp	40
sVEGFR1-RP sVEGFR1 Probe U01134	2388R 2257	5'-TCCGAGCCTGAAAGTTAGCAA-3' 5'-TCCAAATTTAAAGCACAGGAATGATTGTACCAC-3'		
VEGFR2-FP	791F	5'-CTTCGAAGCATCAGCATAAGAAACT-3'	156 bp	40
VEGFR2-RP VEGFR2 Probe AF063658	946R 820	5'-TGGTCATCAGCCCACTGGAT-3' 5'-AACCGAGACCTAAAAACCCAGTCTGGGAGT-3'		
NRP1-FP	1831F	5'-CACAGTGGAAACAGGTGATGACTTC-3'	112 bp	40
NRP1-RP	1942R	5'-AACCATATGTTGGAAACTCTGATTGT-3'		
NRP1 Probe XM_034725	1883	5'-CCACAGAAAAGCCCAAGGTATAGACA-3'		

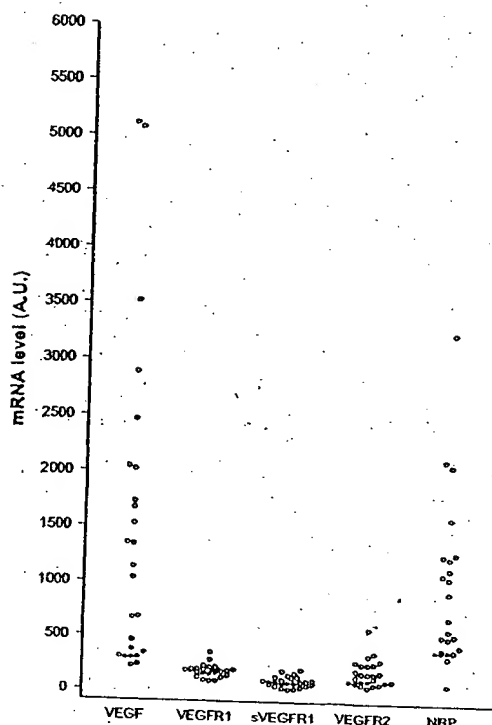


FIGURE 1 – VEGF and VEGF receptors mRNA quantification: scatter of the distribution. Normal brain (black spots) and GBMs (white spots) mRNA levels are expressed as normalized values (as described in Material and Methods; A.U. = arbitrary units). Each point represents the mean of 3 separate experiments.

*End point quantitative PCR for MT1-MMP, MMP-2, MMP-9, TIMP-2 mRNA and VEGF mRNA isoforms*

MT1-MMP, MMP-2, MMP-9, TIMP-2 and VEGF mRNA isoforms (VEGF<sub>189</sub>, VEGF<sub>165</sub>, VEGF<sub>145</sub> and VEGF<sub>121</sub>) were measured in 10 ng aliquots of cDNA using Taq polymerase (Takara, Shiga, Japan) and 5 pmol of each primers (Table I). The thermal cycling conditions included 2 min at 95°C for denaturation and then amplification 15 sec at 94°C, 20 sec at 66°C and 20 sec at 72°C (30 sec for VEGF isoforms) with a final incubation 2 min at 72°C. PCR products were resolved on 2% Nusieve 3:1 agarose gels (BioWhittaker, Rockland, MD) and analyzed using a Fluor-S Multimager (Bio-Rad, Hercules, CA) after ethidium bromide staining. Specific mRNA levels were expressed as the ratio of specific transcripts/28S transcripts. Experiments were repeated at least 3 times in duplicate.

*Real-time quantitative PCR for VEGFR-1, sVEGFR-1, VEGFR-2 and NRP1 mRNA*

Real-time quantitative RT-PCR analyses for VEGFR-1, sVEGFR-1, VEGFR-2, NRP1 mRNAs and 18S rRNA were carried out using the ABI PRISM 7700 Sequence Detection System instrument and software (PE Applied Biosystems). The sequences of the PCR primer pairs and fluorogenic probes that were used for each gene are shown in Table I. A standard curve was generated by 5-fold serial dilution of placenta cDNA to cover the range of 50,000–80 ng and was run in duplicate during every experiment. For each experimental sample, the amount of target gene was determined from this standard curve. The relative expression level of the target gene was normalized against 18S rRNA to compen-

sate for variation in the quality of RNA and the amount of input cDNA (as described by the manufacturer PE Applied Biosystems in User Bulletin 2). PCR was carried out with the TaqMan Universal PCR Master Mix (Applied Biosystems) using 5 µl of diluted cDNA (equivalent to 10 ng total RNA), 200 nM of the probe, and 400 nM primers in a 25 µl final reaction mixture. After a 2 min incubation at 50°C to allow for UNG cleavage, AmpliTaq Gold was activated by an incubation for 10 min at 95°C. Each of the 40 PCR cycles consisted of 15 sec of denaturation at 95°C and hybridization of probe and primers for 1 min at 60°C.

To confirm amplification specificity, the PCR products were also examined by subsequent 2% agarose gel electrophoresis. Experiments were repeated at least 3 times in duplicate.

*Immunohistochemistry for VEGF and MT1-MMP*

Sections (4 µm thick) were cut from formalin-fixed, paraffin embedded tumour tissue. They were hydrated through graded alcohols and incubated in H<sub>2</sub>O<sub>2</sub> (0.3% 15 min). Sections were autoclaved for 11 min at 126°C in citrate buffer pH6 for antigen retrieval (Dako, Glostrup, Denmark). For double immunostaining sections were incubated in primary monoclonal Ab anti-MT1-MMP (Ab-4) 1:100 (Oncogene Research Products, San Diego, CA) followed by peroxidase-conjugated EnVision (Dako). Immunoreactivity was visualized with 3,3'-diaminobenzidine (DAB+, Dako). Sections were then incubated with polyclonal Ab anti-VEGF 1:150 (Santa Cruz, Santa Cruz, CA) for 1 hr at room temperature, followed by alkaline phosphatase-conjugated EnVision (Dako). Immunoreactivity for VEGF was visualized with Fast Red chromogenic substrate (Dako). Single immunostaining was also carried out on serial sections using each primary antibody alone with the corresponding enzyme-chromogene combination. Negative controls were obtained by omitting the primary antibodies.

*Gelatin zymography assay*

MMP-2 and MMP-9 activities were quantified by gelatin zymography on 2 normal brains and 18 GBMs. Ten cryosections (10 µm) were homogenized in buffer (0.1 M Tris-HCl pH 8.1, 0.4% Triton X-100) and centrifuged for 20 min at 5,000g. The pellets were discarded. 25 µg of total protein from homogenate supernatants were mixed with non reducing sample buffer (62.5 mM Tris-HCl, pH 6.8; 2% SDS; 10% glycerol; 0.1% bromophenol blue) and electrophoresed directly on 10% SDS-polyacrylamide gels (SDS-PAGE) containing 0.1% gelatin (w/v).<sup>32</sup> After electrophoresis, gels were washed for 1 hr at room temperature in a 2% (v/v) Triton X-100 solution to remove SDS, transferred to a buffer (50 mM Tris-HCl, pH 7.6, containing 10 mM CaCl<sub>2</sub>) and incubated for 18 hr at 37°C. Gels were stained for 30 min with 0.1% (w/v) Coomassie brilliant blue G250 in 45% (v/v) methanol/10% (v/v) acetic acid and destained in 10% (v/v) acetic acid/20% (v/v) methanol. Gels were analyzed with Quantity One software (version 4.2.2, Bio-Rad Laboratories, Hercules, CA) after densitometric scanning of the gels using a Fluor-S Multimager (BioRad).

*Western blot*

MT1-MMP protein levels were analyzed in 2 normal brains and 15 GBMs. Brain extracts (25 µg) were mixed with 1/2 sample buffer [0.25 M Tris (pH6.8), 10% SDS (w/v), 4% sucrose (v/v), 5% β-mercaptoethanol (v/v) and 0.125% bromophenol blue (w/v)] and boiled for 5 min. They were separated on 10% SDS-PAGE gels and transferred to a PVDF filter (NEN, Boston, MA). After blocking with 5% milk (w/v), 0.1% tween 20 (w/v) in PBS for 2 hr at room temperature, membranes were exposed to the primary antibody (10 µg/ml, clone 113-5B7, Ab-4, Oncogene Research Products, San Diego, CA) at 4°C overnight followed by incubation with a horseradish peroxidase-conjugated rabbit anti-mouse antibody (1.3 µg/ml, Dako, Glostrup, Denmark). Signals were detected with an enhanced chemoluminescence (ECL) kit (NEN, Boston, MA). The relative intensities of the immunoreactive bands were analyzed with Quantity One software (version 4.2.2, Bio-Rad).

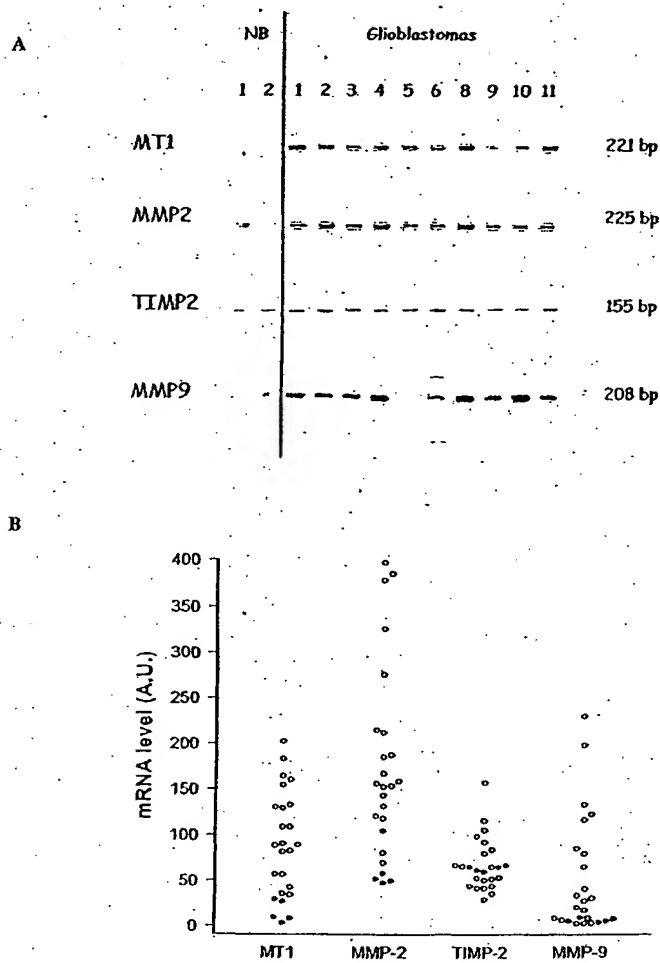


FIGURE 2—MMPs and TIMP-2 mRNA quantification. (a) Representative 2% agarose gels of RT-PCR products for MT1-MMP, MMP-2, TIMP-2 and MMP-9 in 2 normal brains (NB) and 10 GBMs. (b) Scatter plots (as described in Fig. 1). Experiment was repeated at least 3 times in duplicate.

Laboratories) after densitometric scanning of the X-ray films using a Fluor-S Multimager (Bio-Rad).

#### Statistics

VEGF, VEGFRs, MMPs and TIMP-2 expression values in GBMs were correlated using Spearman's test. Correlation was considered significant for 2-tailed  $p$ -value  $< 0.05$ . Statistical analysis was carried out using the Prism 3.0 software (GraphPad, San Diego, CA).

### RESULTS

#### Expression of VEGF and VEGF receptors

VEGF mRNA was present in normal brains (295–375, arbitrary units; mean = 322) and in all GBM samples (217–5,112; mean = 1,774) as reported previously (Fig. 1).<sup>31</sup> In most GBMs, VEGF mRNA levels were raised 2–15-fold above normal brain values. The most abundant isoform in all cases was VEGF<sub>165</sub>, followed by VEGF<sub>121</sub>, VEGF<sub>189</sub> and VEGF<sub>145</sub> (data not shown). VEGFR-1 expression was found at similar levels in GBMs (89–357; mean = 182) and normal controls (154–198; mean = 181). There was no correlation between VEGFR-1 and VEGF mRNA levels ( $p = 0.35$ ) in GBMs. VEGFR-2 was expressed in all GBMs (48–582;

mean = 210) and in 8/20 cases at least twice normal values (87–111; mean = 103). VEGF and VEGFR-2 expressions were correlated significantly ( $p = 0.0035$ ) in GBMs. NRP1 expression varied broadly between GBMs (75–3,260; mean = 1,061) contrasting with a constant baseline expression in normal controls (383–397; mean = 390). In tumours, NRP1 correlated with VEGFR-2 ( $p = 0.0119$ ) but not with VEGF ( $p = 0.084$ ), nor VEGFR-1 ( $p = 0.066$ ). sVEGFR-1 was expressed at low levels both in normal brains (84–92; mean = 87) and GBMs (25–208; mean = 101). sVEGFR-1, however, was found to correlate with VEGFR-1 ( $p = 0.0289$ ), VEGFR-2 ( $p = 0.0029$ ), and NRP1 ( $p = 0.0027$ ) but not with VEGF ( $p = 0.053$ ).

#### Expression of MMPs and TIMP-2

MT1-MMP, MMP-2 and MMP-9 were expressed in both normal brains and GBMs but at much higher levels in the latter (Fig. 2a). MT1-MMP mRNA levels were constantly higher in GBMs (34–202; mean = 106) than in normal controls (3–29; mean = 15). MMP-2 and MMP-9 mRNA levels were higher than controls in 18/20 and 14/20 cases respectively (Fig. 2b) and correlated with each other ( $p = 0.0187$ ). MT1-MMP mRNA levels correlated with MMP-2 ( $p = 0.0008$ ) and MMP-9 ( $p = 0.005$ ). TIMP-2 had a non-discriminative distribution in relation to the controls. TIMP-2

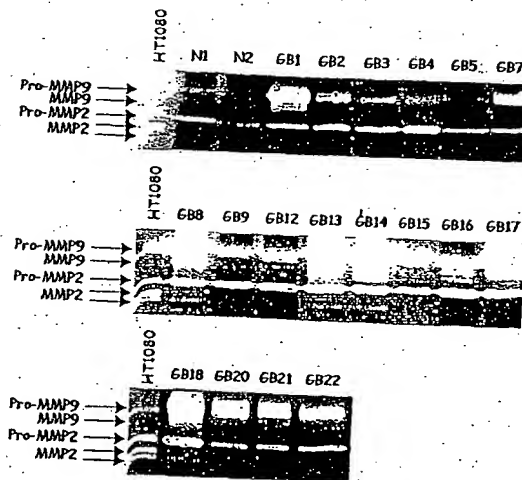


FIGURE 3—Zymographic analysis of MMP-2 and MMP-9 in tissue extracts either from normal brain (N) or GBMs (GB). Medium conditioned by human HT1080 cells was included as positive control. Positions of pro-MMP-9, MMP-9, pro-MMP-2 and MMP-2 are indicated by arrows.

TABLE II—ZYMOGRAMS AND WESTERN BLOT QUANTIFICATION OF MMPs IN NORMAL BRAIN AND GBMs<sup>1</sup>

	Gelatin zymography				Western blot MT1-MMP
	proMMP-9	MMP-9	proMMP-2	MMP-2	
N1	0.41	0.00	1.52	0.00	0.44
N2	0.28	0.00	0.83	0.00	0.47
GB1	9.00	1.25	2.45	0.35	1.58
GB2	4.28	0.00	2.88	0.09	0.47
GB3	1.67	0.00	2.86	0.00	0
GB4	1.40	0.00	4.33	0.03	0
GB5	0.47	0.00	1.15	0.00	ND
GB7	5.29	0.00	1.12	0.00	0
GB8	5.75	3.78	4.63	0.65	0.55
GB9	5.02	0.00	4.09	0.00	0.45
GB12	5.08	0.00	4.79	0.20	3.18
GB13	6.86	4.58	4.93	0.48	1.55
GB14	5.59	6.75	5.53	0.76	0.94
GB15	8.00	0.00	5.13	0.00	ND
GB16	5.55	1.10	5.66	0.00	2.50
GB17	6.61	2.04	6.00	0.00	0
GB18	14.79	4.34	8.20	0.40	1.97
GB20	18.85	0.00	2.12	0.00	0
GB21	16.52	0.00	2.37	0.00	ND
GB22	20.66	0.00	5.11	0.00	0.86

<sup>1</sup>Extracts expressed as arbitrary units. N, normal brain; GB, glioblastoma; ND, not determined.

was correlated with MT1-MMP ( $p = 0.0019$ ) and MMP-2 ( $p = 0.0002$ ) but not with MMP-9 ( $p = 0.1408$ ).

#### Correlation between MT1-MMP protein and activated MMP-2 and -MMP-9

By gelatin zymography, pro-MMP-2 and pro-MMP-9 were detected in the 18 GBMs and 2 controls examined (Fig. 3, Table II). In most GBMs, levels of these inactive forms were higher than in normal brains and were correlated with their respective mRNA levels (MMP-2:  $p < 0.0001$ ; MMP-9  $p = 0.01$ ). Activated forms of MMP-2 and MMP-9 were not found in normal brain. By contrast, they were present in 8/18 (MMP-2) and 7/18 (MMP-9) GBMs. MT1-MMP protein levels were quantified by Western blot in 15 GBMs and 2 normal brains (Fig. 4, Table II). They were signifi-

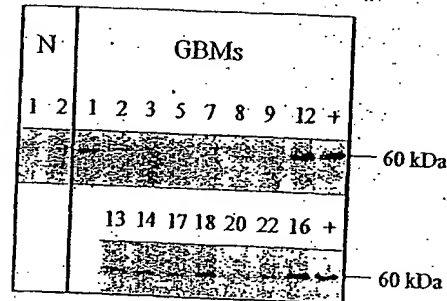


FIGURE 4—Western blot analysis using the ab4 antibody (clone 113-5B7) raised against the catalytic domain of MT1-MMP. Protein extracts from MT1-MMP transfected A2058 cells (clone SL5,<sup>28</sup>) were used as a positive control (+). MT1-MMP protein is detected in normal brain (N) and GBMs.

TABLE III—CORRELATION BETWEEN VEGF, VEGF RECEPTORS, MMPs AND TIMP-2 EXPRESSED AS  $p$  VALUES DERIVED FROM SPEARMAN'S TEST<sup>1</sup>

	MT1-MMP	MMP-2	MMP-9	TIMP-2
VEGF	0.0250	0.0245	0.0053	0.0094
VEGFR-1	0.0073	0.0710	<0.0001	0.4542
VEGFR-2	<0.0001	0.0168	0.0004	0.0153
NRP1	0.0053	0.1334	0.2457	0.5480
sVEGFR-1	0.0313	0.0469	0.0194	0.2563

cantly correlated with zymogram-derived activated MMP-2 levels ( $p = 0.0226$ ) but not with activated MMP-9 levels ( $p = 0.06$ ). Interestingly, MT1-MMP protein and mRNA levels were correlated significantly ( $p = 0.089$ ), arguing for a predominantly transcriptional regulation in GBMs.

#### Correlation between VEGF network and MMPs

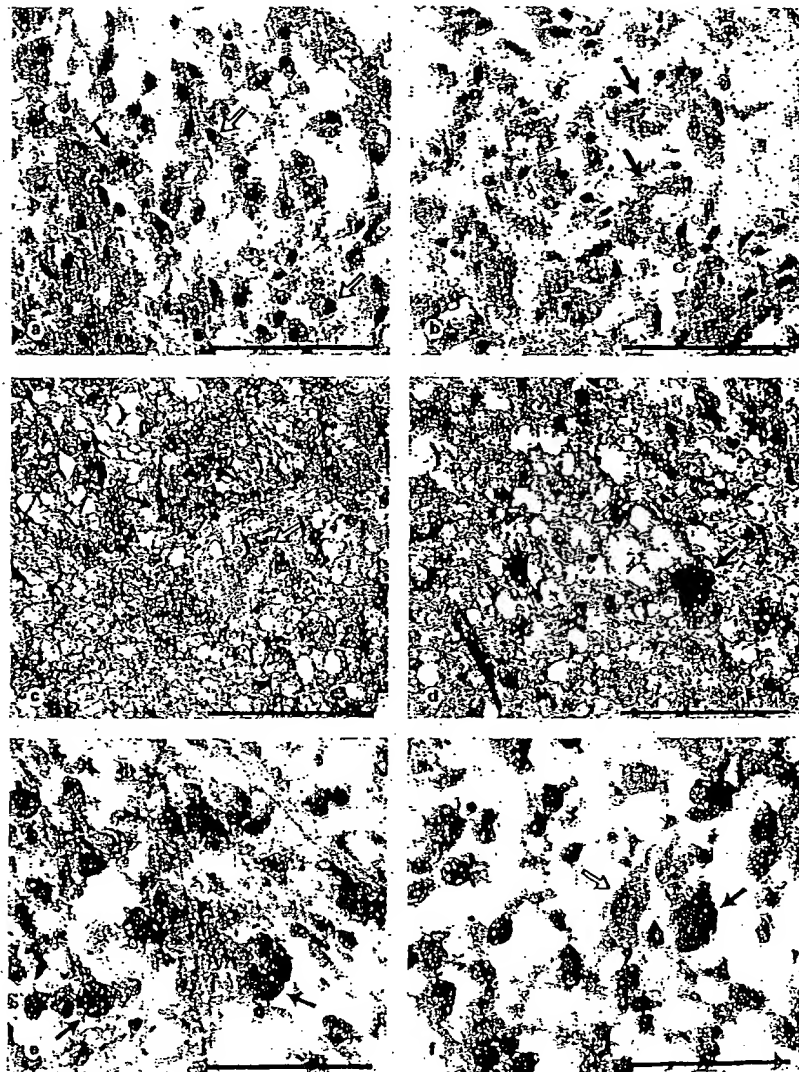
mRNA levels of VEGF and VEGF receptors were compared to MT1-MMP, MMP-2, MMP-9 and TIMP-2 (Table III). There was a significant correlation between VEGF expression and MT1-MMP, MMP-2 and MMP-9. A similar correlation was also observed between VEGFR-2 and MMPs. Interestingly, TIMP-2 expression was correlated with VEGF and VEGFR-2 but not with other VEGF receptors.

#### Immunohistochemistry for VEGF and MT1-MMP

VEGF immunoreactivity was shown in both tumour and endothelial cells, as previously reported (Fig. 5a,b).<sup>31</sup> By single immunostaining, MT1-MMP was detected in glioblastoma cells as a diffuse cytoplasmic staining (Fig. 5c,d). MT1-MMP positivity was also seen in endothelial cells and perivascular cells (Fig. 5c). By double immunostaining, we observed the co-localization of VEGF and MT1-MMP in the cytoplasm of numerous tumour cells (Fig. 5e,f).

#### DISCUSSION

GBMs are highly malignant tumours with poor prognosis. They show major microvascular proliferation and express high levels of VEGF.<sup>4-9</sup> VEGF is a strong mitogen for endothelial cells thereby promoting angiogenesis. Previous reports<sup>23,24</sup> have suggested that VEGF also stimulates tumour cell invasion, migration and survival in malignant epithelial cells through an autocrine loop by which overexpression of MMPs induces VEGF secretion and leads to subsequent amplification of cell proliferation and protection against apoptosis. We and others reported previously that in human melanoma and breast carcinoma cells, MT1-MMP upregulates VEGF expression whereas TIMP-2 reduces it.<sup>28,29,35</sup> Therefore the



**FIGURE 5**—Immunohistochemistry (scale bar = 50  $\mu$ m). (a,b) VEGF positive tumour and endothelial cells (plain arrows) show granular red staining of the cytoplasm. Negative cells (empty arrow) are seen in their close vicinity. (c) MT1-MMP positive cells show a strong brown cytoplasmic staining. They include tumour cells (plain arrow) and endothelial cells (empty arrow). (d) MT1-MMP positive tumour cells (plain arrow) are mixed with negative cells (empty arrow). (e,f) Double staining with VEGF (red) and MT1-MMP (brown). Double positive tumour cells (plain arrows) contrast with negative or single weakly positive cell (empty arrow).

pericellular proteolysis mediated by MT1-MMP in GBMs could also induce an autocrine loop resulting in enhanced VEGF expression. In turn, VEGF could act as a paracrine factor on endothelial cells to stimulate angiogenesis or possibly as an autocrine factor promoting glioblastoma cells survival-migration and invasion as demonstrated recently in the various tumour cell culture models.

We compared the expression of VEGF and its receptors with MT1-MMP, MMP-2 and MMP-9 in 20 GBMs and 5 normal brains. The expression of these MMPs was markedly increased in most GBMs with excellent correlation between mRNA and protein levels. MT1-MMP expression has been shown previously to correlate with glioma aggressiveness and its transfection in different tumour cell lines triggers an angiogenic phenotype and promotes tumour growth.<sup>20,28–30,36–38</sup> A majority of GBMs (17/20) also expressed high levels of VEGF, as previously reported, with a strong correlation between VEGF and MT1-MMP gene expression levels. Double immunostaining studies showed co-expression of VEGF and MT1-MMP by the same tumour cells. These data suggest that the transcrip-

tional control of VEGF by MT1-MMP could be operative not only *in vitro* but also *in vivo* in human GBMs.

MT1-MMP could also promote the growth of GBMs by its ability to activate MMP-2 in the presence of low concentration of TIMP-2.<sup>40</sup> Pro-MMP-2 activation occurs after the formation of a ternary complex that contains pro-MMP-2 linked to cell surface MT1-MMP via a TIMP-2 bridge. In accordance with this hypothesis, we found that MMP-2 activation occurred in 8/18 of our GBMs<sup>20,41</sup> among which 7/7 tested for MT1-MMP showed high contents of this protease.

Activated MMP-9 was also found in 7/18 of our GBMs. This is an interesting finding as active MMP-9 is able to mobilize VEGF from its ECM reservoir.<sup>39</sup> Therefore, MMPs could promote VEGF-mediated angiogenesis in GBMs by both transcriptional (MT1-MMP) and post translational (MMP-9) mechanisms.

VEGF binding to VEGFR-2 triggers the proliferation and migration of endothelial cells whereas its binding to VEGFR-1 has



opposite effects on glioblastoma cell lines.<sup>12,42</sup> In our study, VEGF mRNA levels were correlated with VEGFR-2 but not VEGFR-1, NRP1 and sVEGFR-1. Collectively our data suggest that GBMs display a specific and complex pattern of VEGF receptors, transducing VEGF signaling toward cell proliferation and migration.

In conclusion, our study adds to the evidence for an interplay between metalloproteinases and VEGF in human GBMs as previously documented in experimental tumours. Because of its dual ability to activate MMP-2 and to up-regulate VEGF, MT1-MMP might be of central importance in the growth of human glioblastomas and represent an interesting target for anti-cancer treatments.

#### ACKNOWLEDGEMENTS

The authors wish to express their gratitude to the neurosurgeons of the CHU-Liège (Prof. A. Stevenaert), and CHR-Liège (Prof. J.

Born) and Dr. A. Michone (VUB-Brussels) for their contribution to the collection of cases, to Dr. L. de Leval for her assistance in immunohistochemistry, and to M.R. Pignon for her excellent technical assistance. This work was supported by grants from the Communauté Française de Belgique (Actions de Recherches Concertées), the Commission of European Communities, the Fonds de la Recherche Scientifique Médicale, the Fonds National de la Recherche Scientifique (FNRS, Belgium), the Fédération Belge Contre le Cancer, the Fonds spéciaux de la Recherche (University of Liège), the Centre Anticancéreux près l'Université de Liège, the FB Assurances, the Fondation Léon Frédéricq (University of Liège), the D.G.T.R.E. from the "Région Wallonne", the Fonds d'Investissements de la Recherche Scientifique (CHU, Liège, Belgium), the Interuniversity Attraction Poles (I.U.A.P.) from the Federal Office for Scientific, Technical and Cultural Affairs (O.S.T.C., Brussels, Belgium).

#### REFERENCES

- Folkman J, Klagsbrun M. Angiogenic factors. *Science* 1987;235:442-7.
- Carmeliet P, Jain RK. Angiogenesis in cancer and other diseases. *Nature* 2000;407:249-57.
- Plate KH, Breier G, Weich HA, Risau W. Vascular endothelial growth factor is a potential tumour angiogenesis factor in human gliomas in vivo. *Nature* 1992;359:845-8.
- Oehring RD, Miletic M, Valtier MM, Pietsch T, Neumann J, Fimmers R, Schlegel U. Vascular endothelial growth factor (VEGF) in astrocytic gliomas—a prognostic factor? *J Neurooncol* 1999;45:117-25.
- Pietsch T, Valtier MM, Wolf HK, von Deimling A, Huang HJ, Cavenee WK, Wiestler OD. Expression and distribution of vascular endothelial growth factor protein in human brain tumors. *Acta Neuropathol (Berl)* 1997;93:109-17.
- Nishikawa R, Cheng SY, Nagashima R, Huang HJ, Cavenee WK, Matsutani M. Expression of vascular endothelial growth factor in human brain tumors. *Acta Neuropathol (Berl)* 1998;96:453-62.
- Kleihues P, Cavenee WK. Pathology and genetics of tumours of the nervous system. Lyon: IARC Press, 2000. 29-39.
- Plate KH, Breier G, Weich HA, Mennel HD, Risau W. Vascular endothelial growth factor and glioma angiogenesis: coordinate induction of VEGF receptors, distribution of VEGF protein and possible in vivo regulatory mechanisms. *Int J Cancer* 1994;59:520-9.
- Machein MR, Plate KH. VEGF in brain tumors. *J Neurooncol* 2000;50:109-20.
- Ding H, Wu X, Roncari L, Lau N, Shannon P, Nagy A, Guha A. Expression and regulation of neuropilin-1 in human astrocytomas. *Int J Cancer* 2000;88:584-92.
- Kunkel P, Ulbricht U, Bohlen P, Brockmann MA, Fillbrandt R, Stavrou D, Westphal M, Lamszus K. Inhibition of glioma angiogenesis and growth in vivo by systemic treatment with a monoclonal antibody against vascular endothelial growth factor receptor-2. *Cancer Res* 2001;61:6624-8.
- Herold-Mende C, Steiner HH, Andl T, Riede D, Buttler A, Reisser C, Fussenig NE, Mueller MM. Expression and functional significance of vascular endothelial growth factor receptors in human tumor cells. *Lab Invest* 1999;79:1573-82.
- Hiratsuka S, Minowa O, Kuno J, Noda T, Shibuya M. Flt-1 lacking the tyrosine kinase domain is sufficient for normal development and angiogenesis in mice. *Proc Natl Acad Sci USA* 1998;95:9349-54.
- Soker S, Takashima S, Miao HQ, Neufeld G, Klagsbrun M. Neuropilin-1 is expressed by endothelial and tumor cells as an isoform-specific receptor for vascular endothelial growth factor. *Cell* 1998;92:735-45.
- Egeblad M, Werb Z. New functions for the matrix metalloproteinases in cancer progression. *Nat Rev Cancer* 2002;2:161-74.
- Pepper MS. Role of the matrix metalloproteinase and plasminogen activator-plasmin systems in angiogenesis. *Arterioscler Thromb Vasc Biol* 2001;21:1104-17.
- Sternlicht MD, Werb Z. How matrix metalloproteinases regulate cell behavior. *Annu Rev Cell Dev Biol* 2001;17:463-516.
- Sounni NE, Jonssen M, Foidart JM, Noel A. Membrane Type-1 Matrix Metalloproteinase and TIMP2 in tumor angiogenesis. *Matrix Biol* 2003;22:55-61.
- Belien AT, Paganetti PA, Schwab ME. Membrane-type 1 matrix metalloproteinase (MT1-MMP) enables invasive migration of glioma cells in central nervous system white matter. *J Cell Biol* 1999;144:373-84.
- Lampert K, Machein U, Machein MR, Conca W, Peter HH, Volk B. Expression of matrix metalloproteinases and their tissue inhibitors in human brain tumors. *Am J Pathol* 1998;153:429-37.
- Sato H, Takino T, Okada Y, Cao J, Shinagawa A, Yamamoto E, Seiki M. A matrix metalloproteinase expressed on the surface of invasive tumour cells. *Nature* 1994;370:61-5.
- Knauper V, Will H, Lopez-Otin C, Smith B, Atkinson SJ, Stanton H, Hembry RM, Murphy G. Cellular mechanisms for human procollagenase-3 (MMP-13) activation. Evidence that MT1-MMP (MMP-14) and gelatinase a (MMP-2) are able to generate active enzyme. *J Biol Chem* 1996;271:17124-31.
- Koshikawa N, Giannelli G, Cirulli V, Miyazaki K, Quaranta V. Role of cell surface metalloproteinase MT1-MMP in epithelial cell migration over laminin-5. *J Cell Biol* 2000;148:615-24.
- Gilles C, Polette M, Coraux C, Tournier JM, Meneguzzi G, Munaut C, Volders L, Rousselle P, Birembaut P, Foidart JM. Contribution of MT1-MMP and of human laminin-5 gamma2 chain degradation to mammary epithelial cell migration. *J Cell Sci* 2001;114:2967-76.
- Kajita M, Itoh Y, Chiba T, Mori H, Okada A, Kinoh H, Seiki M. Membrane-type 1 matrix metalloproteinase cleaves CD44 and promotes cell migration. *J Cell Biol* 2001;153:893-904.
- Deryugina EI, Ratnikov B, Monosov E, Postnova TI, DiScipio R, Smith JW, Strongin AY. MT1-MMP initiates activation of pro-MMP-2 and integrin  $\alpha$ 3 promotes maturation of MMP-2 in breast carcinoma cells. *Exp Cell Res* 2001;263:209-23.
- Zhou Z, Apte SS, Soininen R, Cao R, Baakkhi GY, Rausser RW, Wang J, Cao Y, Tryggvason K. Impaired endochondral ossification and angiogenesis in mice deficient in membrane-type matrix metalloproteinase 1. *Proc Natl Acad Sci USA* 2000;97:4052-7.
- Sounni NE, Baramova EN, Munaut C, Maquie E, Frankenne F, Foidart JM, Noel A. Expression of membrane type 1 matrix metalloproteinase (MT1-MMP) in A2058 melanoma cells is associated with MMP-2 activation and increased tumor growth and vascularization. *Int J Cancer* 2002;98:23-8.
- Sounni NE, Devy L, Hajitou A, Frankenne F, Munaut C, Gilles C, Deroanne C, Thompson EW, Foidart JM, Noel A. MT1-MMP expression promotes tumor growth and angiogenesis through an upregulation of vascular endothelial growth factor expression. *FASEB J* 2002;16:555-64.
- Deryugina EI, Sorocanu L, Strongin AY. Upregulation of vascular endothelial growth factor by membrane-type 1 matrix metalloproteinase stimulates human glioma xenograft growth and angiogenesis. *Cancer Res* 2002;62:580-8.
- Munaut C, Boniver J, Foidart JM, Deprez M. Macrophage migration inhibitory factor (MIF) expression in human glioblastomas correlates with VEGF expression. *Neuropathol Appl Neurobiol* 2002;28:452-60.
- Munaut C, Noel A, Weidle UH, Krell HW, Foidart JM. Modulation of the expression of interstitial and type-IV collagenases in coculture of HT1080 fibrosarcoma cells and fibroblasts. *Invasion Metastasis* 1995;15:169-78.
- Bachelder RE, Crago A, Chung J, Wendt MA, Shaw LM, Robinson G, Mercurio AM. Vascular endothelial growth factor is an autocrine survival factor for neuropilin-expressing breast carcinoma cells. *Cancer Res* 2001;61:5736-40.
- Chung J, Bachelder RE, Lipscomb EA, Shaw LM, Mercurio AM. Integrin  $\alpha$ 6 $\beta$ 4 regulation of eIF-4E activity and VEGF translation: a survival mechanism for carcinoma cells. *J Cell Biol* 2002;158:165-74.
- Hajitou A, Sounni NE, Devy L, Grigori-Debrus C, Lewalle JM, Li H, Deroanne CF, Lu H, Colige A, Nusgens BV, Frankenne F, Maron A.

- et al. Downregulation of vascular endothelial growth factor by tissue inhibitor of metalloproteinase-2: effect on in vivo mammary tumor growth and angiogenesis. *Cancer Res* 2001;61:3450-7.
36. Forsyth PA, Laing TD, Gibson AW, Rewcastle NB, Brasher P, Sutherland G, Johnston RN, Edwards DR. High levels of gelatinase-B and active gelatinase-A in metastatic glioblastoma. *J Neurooncol* 1998;36:21-9.
37. Nakada M, Nakamura H, Ikeda E, Fujimoto N, Yamashita J, Sato H, Seiki M, Okada Y. Expression and tissue localization of membrane-type 1, 2, and 3 matrix metalloproteinases in human astrocytic tumors. *Am J Pathol* 1999;154:417-28.
38. Kachra Z, Beaulieu E, Delbecchi L, Mousseau N, Berthelet F, Moun-djian R, Del Maestro R, Beliveau R. Expression of matrix metalloproteinases and their inhibitors in human brain tumors. *Clin Exp Metastasis* 1999;17:555-66.
39. Bergers G, Brekken R, McMahon G, Vu TH, Itoh T, Tamaki K, Tanzawa K, Thorpe P, Itohara S, Werb Z, Hanahan D. Matrix metalloproteinase-9 triggers the angiogenic switch during carcinogenesis. *Nat Cell Biol* 2000;2:737-44.
40. Strongin AY, Collier I, Bannikov G, Marmer BL, Grant GA, Goldberg GL. Mechanism of cell surface activation of 72-kDa type IV collagenase. Isolation of the activated form of the membrane metalloproteinase. *J Biol Chem* 1995;270:5331-8.
41. Nakada M, Kita D, Futami K, Yamashita J, Fujimoto N, Sato H, Okada Y. Roles of membrane type 1 matrix metalloproteinase and tissue inhibitor of metalloproteinases 2 in invasion and dissemination of human malignant glioma. *J Neurosurg* 2001;94:464-73.
42. Millauer B, Witzmann-Voos S, Schmurch H, Martinez R, Moller NP, Risau W, Ullrich A. High affinity VEGF binding and developmental expression suggest Flk-1 as a major regulator of vasculogenesis and angiogenesis. *Cell* 1993;72:835-46.

INTERNATIONAL

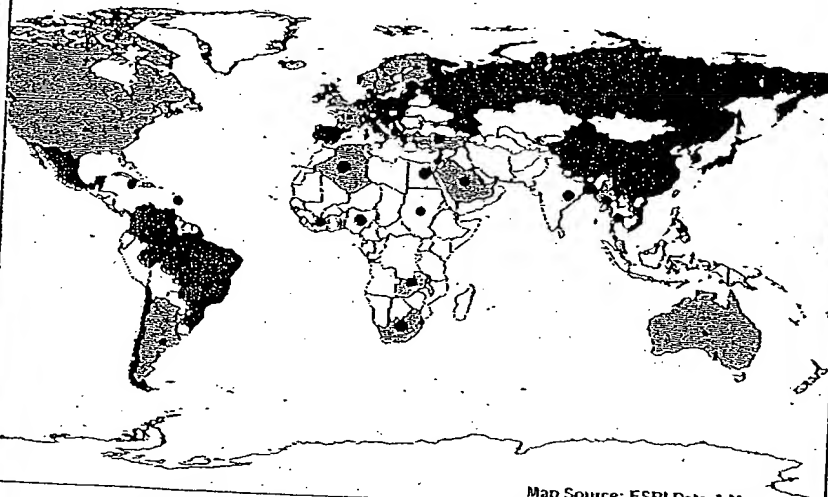
# Journal of Cancer

HEALTH SCIENCES LIBRARY  
UNIVERSITY OF WISCONSIN

SEP 10 2003

1305 Linden Drive  
Madison, WI 53706

## *Helicobacter pylori* and gastric cancer: facing the enigmas



Map Source: ESRI Data & Maps

Prevalence of *H. pylori*  
(%)

- 1st quartile (19.39.9)
- 2nd quartile (39.9-67.5)
- 3rd quartile (67.5-81.2)
- 4th quartile (81.2-96.3)

Gastric cancer incidence \*  
(100 000)

- ≤ 5
- ▨ 5-10
- ▩ 10-15
- 15-20
- > 20

\* Age and sex-standardized



## International Union Against Cancer

### COUNCIL

J. SEFFREN, President (USA)  
L.J. DENIS, Treasurer (Belgium)  
J. BAITY, Chair, Finance Committee (USA)

G. BRIEN (Australia)  
R.C. BURTON (Australia)  
M. DAUBE (Australia)  
K.A. DINSHAW (India)  
L. ELOVAINIO (Finland)  
M.K. GOSPODAROWICZ (Canada)  
R.T. HUDSON (Ireland)  
T. KITAGAWA (Japan)  
A. PUNDALIK KURKURE (India)  
R.E. LENHARD (USA)  
A. LLOMBART-BOSCH (Italy)

M. LUWIA (Indonesia)  
C. MALLINSON (UK)  
I.H. MARCHESI (Brazil)  
H.F. MICKELSON (USA)  
K. NILSSON (Sweden)  
S. OMAR (Egypt)  
T. PHILIP (France)  
E. ROBINSON, Past President (Israel)  
Y. SALOOJEE (South Africa)  
H. SANCHE-GARNIER (France)

R.J. SCHWEITZER (USA)  
O. SOREIDE (Norway)  
K. TAJIMA (Japan)  
W. WEBER (Switzerland)  
S. WILKINSON (UK)  
D. ZACKS (USA)  
D. ZARIDZE (Russia)  
Y. HUI ZHANG (China)  
M. ZIV (Israel)  
M. ZUR HAUSEN (Germany)

UICC, URL: <http://www.uicc.org>  
UICC, E-mail: [info@uicc.org](mailto:info@uicc.org)

The International Union Against Cancer (UICC) is devoted exclusively to all aspects of the world-wide fight against cancer. Its objectives are to advance scientific and medical knowledge in research, diagnosis, treatment and prevention of cancer, and to promote all other aspects of the campaign against cancer throughout the world. Particular emphasis is placed on professional and public education.

Founded in 1933, the UICC is a non-governmental, independent association of more than 290 member organizations in over 80 countries. Members are voluntary cancer leagues and societies, cancer research and/or treatment centers, and in some countries ministries of health.

The UICC is non-profit, non-political, and non-sectarian. Its headquarters are in Geneva, Switzerland. It creates and carries out programs around the world in collaboration with hundreds of volunteer experts. Supported by membership dues, national subscriptions, grants and donations, its annual budget is about US \$4 million.

The UICC is governed by its members which meet in General Assembly every 4 years. Its elected Council and Executive Committee are responsible for Program structure and implementation.

The UICC organizes an International Cancer Congress every 4 years as well as annual symposia, workshops, and training courses. It publishes the *International Journal of Cancer* (30 issues per year), *UICC News* (quarterly), the *International Calendar of Meetings on Cancer* (bi-annually), and a number of technical reports, textbooks and manuals.

### SUBSCRIPTION INFORMATION

© 2003 Wiley-Liss, Inc., a Wiley Company. All rights reserved. No part of this publication may be reproduced in any form or by any means, except as permitted under section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the publisher, or authorization through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, telephone: (978) 750-8400, fax: (978) 750-4470. Requests to the publisher for permission should be addressed to the Permissions Department, c/o John Wiley & Sons, Inc., 111 River St., Hoboken, NJ 07030. Fax: (201) 748-6008; Tel.: (201) 748-6011; <http://www.wiley.com/go/permissions>.

*International Journal of Cancer* (Print ISSN 0020-7136; Online ISSN 1097-0215) is published 30 times a year, semi-monthly with extra issues in January, March, May, July, September, and November, by Wiley-Liss, Inc., through Wiley Subscription Services, Inc., a Wiley Company. Send subscription inquiries in care of John Wiley & Sons, Inc., Attn: Journals Admin Dept UK, 111 River St., Hoboken, NJ 07030, (201) 748-6645.

Advertising inquiries should be addressed to Advertising Department, c/o John Wiley & Sons, Inc., 111 River St., Hoboken, NJ 07030. Telephone: (201) 748-6921.

Offprint sales and inquiries should be directed to the Customer Service Department, in care of John Wiley & Sons, 111 River St., Hoboken, NJ 07030. Telephone: (201) 748-8776.

Subscription price: Volumes 103-107, 2003 (30 issues). Print only: \$2,495 worldwide.

Electronic only: \$2,495 worldwide. A combination price of \$2,620 worldwide includes the subscription in both electronic and print formats. All subscriptions containing a print element, shipped outside the U.S., will be sent by air. Payment must be made in U.S. dollars drawn on U.S. bank. Periodicals postage paid at Hoboken, NJ and at additional mailing offices. Postmaster: send address changes to INTERNATIONAL JOURNAL OF CANCER, Subscription Distribution, c/o John Wiley & Sons, Inc., 111 River St., Hoboken, NJ 07030.

Change of address: Please forward to the subscriptions address listed above 6 weeks prior to move, enclose present mailing label with change of address. Claims for missing issues: Claims for undelivered copies will be accepted only after the following issue has been received. Please enclose a mailing label or cite your subscriber reference number. Missing copies will be supplied when losses have been sustained in transit and where reserve stock permits. Send claims in care of John Wiley & Sons, Inc., Attn: Journals Admin Dept UK, 111 River St., Hoboken, NJ 07030.

Indexed by: EMBASE/Excerpta Medica • Current Contents/Life Sciences • Science Citation Index • Scisearch • BIOSIS Data Base • Index Medicus • Cambridge Scientific Abstracts • Chemical Abstracts • Reference Update • Smoking and Health Database.

This journal is printed on acid-free paper.

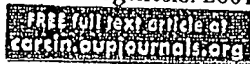
### INTERNATIONAL JOURNAL OF CANCER

2003 Wiley-Liss, Inc.

The *International Journal of Cancer* is published for the International Union Against Cancer by Wiley-Liss, Inc., a division of John Wiley & Sons, Inc. Five volumes are issued annually, each consisting of six numbers.

Abstracting and other journals may reprint the summaries of articles without requesting authorization. Authors alone are responsible for views expressed in signed articles. The mention of specific companies or of certain manufacturers' products does not imply that they are endorsed or recommended by the International Union Against Cancer.

References: When quoting from the *International Journal of Cancer*, please use the official abbreviation: *Int. J. Cancer*



**DNA hypermethylation is a mechanism for loss of expression of the HLA class I genes in human esophageal squamous cell carcinomas.**

Nie Y, Yang G, Song Y, Zhao X, So C, Liao J, Wang LD, Yang CS.

Laboratory for Cancer Research, College of Pharmacy, Rutgers-The State University of New Jersey, 164 Frelinghuysen Road, Piscataway, NJ 08854-8020, USA.

The three human leukocyte antigen (HLA) class I antigens, HLA-A, HLA-B and HLA-C, play important roles in the elimination of transformed cells by cytotoxic T cells. Frequent loss of expression of these antigens at the cell surface has been observed in many human cancers. Various mechanisms for post-transcriptional regulation have been proposed and tested but the molecular mechanisms for transcriptional regulation are not clear. We show by immunohistochemistry that the HLA class I antigens are absent in 26 of 29 (89%) samples of human esophageal squamous cell carcinomas (ESCC). Eleven of the 26 ESCC samples lost mRNA expression for at least one of the HLA genes, as shown by RT-PCR. DNA from the 29 pairs of ESCC and neighboring normal epithelium were examined for CpG island hypermethylation, homozygous deletion, microsatellite instability (MSI) and loss of heterozygosity (LOH). DNA from normal epithelial tissues had no detectable methylation of the CpG islands of any of these gene loci. Thirteen of 29 ESCC samples (45%) exhibited methylation of one or more of the three HLA loci and six samples (21%) exhibited methylation of all three loci. The HLA-B gene locus was most frequently methylated (38%). HLA-B mRNA expression in an ESCC cell line, where HLA-B was hypermethylated and did not express mRNA, was activated after treatment with 5-aza-2'-deoxycytidine. Homozygous deletion of these three gene loci was not observed. Relatively low rates of LOH and MSI were observed for the microsatellite markers D6S306, D6S258, D6S273 and D6S1666, close to the HLA-A, -B and -C loci, although a high ratio of LOH was observed at a nearby locus (represented by the markers D6S1051 and D6S1560), where the tumor suppressor gene p21(Waf1) resides. A strong correlation between genetic alterations and mRNA inactivation was observed in the ESCC samples. Our results indicate that HLA class I gene expression was frequently down-regulated in ESCC at both the protein and mRNA levels and that hypermethylation of the promoter regions of the HLA-A, -B and -C genes is a major mechanism of transcriptional inactivation.

PMID: 11577000 [PubMed - indexed for MEDLINE]

95: Hum Pathol. 2003 Jul;34(7):639-45.

Related Articles, Links

Comment in:

- Hum Pathol. 2003 Jul;34(7):635-8.

Human Pathology

**Molecular and immunohistochemical analysis of HER2/neu oncogene in synovial sarcoma.**

**Nuciforo PG, Pellegrini C, Fasani R, Maggioni M, Coggi G, Parafloriti A, Bosari S.**

Department of Medicine, Surgery and Dental Sciences, University of Milan, A.O.S. Paolo and IRCCS Ospedale Maggiore, Italy.

Amplification and/or overexpression of HER2/neu have been documented in many types of epithelial tumor and recently has been reported in sarcomas, particularly in osteosarcomas. But the role of HER2/neu alterations in soft tissue tumors remains poorly understood. Thus the present study investigates the expression of HER2/neu in 13 patients with synovial sarcoma (SS). In this study, HER2/neu mRNA levels were measured in frozen tissue samples using a real-time reverse transcription-polymerase chain reaction assay; protein expression was assessed by immunohistochemistry using an anti-HER2/neu polyclonal antibody. Six normal skeletal muscle specimens were used to establish basal levels of HER2/neu mRNA. HER2/neu transcripts were detected in all normal tissues and SSs. Four of 13 sarcomas (31%) demonstrated HER2/neu mRNA levels above the mean value, whereas 3 tumors (23%) displayed HER2/neu protein overexpression. Both membranous and cytoplasmic patterns of immunostaining were observed, and a strong correlation was found between protein expression and mRNA level ( $P = 0.01$ ). Increased HER2/neu mRNA levels were significantly associated with a lower risk of developing recurrences ( $P = 0.02$ ). Moreover, none of the patients with HER2/neu overexpression developed metastasis. Our data demonstrate that HER2/neu is expressed in SSs and that both membrane and cytoplasmic HER2/neu expression correlate with mRNA levels. Our results show that the presence of increased levels of HER2/neu in SSs is associated with a more favorable clinical course. Further studies are needed to assess the role of this oncogene in SSs and to evaluate the application of inhibitory humanized monoclonal antibodies in the treatment regimens for this malignancy.

PMID: 12874758 [PubMed - indexed for MEDLINE]

## Original Contributions

### Molecular and Immunohistochemical Analysis of HER2/neu Oncogene in Synovial Sarcoma

PAOLO GIOVANNI NUCIFORO, MD, CATERINA PELLEGRINI, PhD,  
ROBERTA FASANI, MD, MARCO MAGGIONI, MD,  
GUIDO COGGI, MD, ANTONINA PARAFIORITI, MD,  
AND SILVANO BOSARI, MD

Amplification and/or overexpression of HER2/neu have been documented in many types of epithelial tumor and recently has been reported in sarcomas, particularly in osteosarcomas. But the role of HER2/neu alterations in soft tissue tumors remains poorly understood. Thus the present study investigates the expression of HER2/neu in 13 patients with synovial sarcoma (SS). In this study, HER2/neu mRNA levels were measured in frozen tissue samples using a real-time reverse transcription-polymerase chain reaction assay; protein expression was assessed by immunohistochemistry using an anti-HER2/neu polyclonal antibody. Six normal skeletal muscle specimens were used to establish basal levels of HER2/neu mRNA. HER2/neu transcripts were detected in all normal tissues and SSs. Four of 13 sarcomas (31%) demonstrated HER2/neu mRNA levels above the mean value, whereas 3 tumors (23%) displayed HER2/neu protein overexpression. Both membranous and cytoplasmic patterns of immunostaining were observed, and a strong correlation was

found between protein expression and mRNA level ( $P = 0.01$ ). Increased HER2/neu mRNA levels were significantly associated with a lower risk of developing recurrences ( $P = 0.02$ ). Moreover, none of the patients with HER2/neu overexpression developed metastasis. Our data demonstrate that HER2/neu is expressed in SSs and that both membrane and cytoplasmic HER2/neu expression correlate with mRNA levels. Our results show that the presence of increased levels of HER2/neu in SSs is associated with a more favorable clinical course. Further studies are needed to assess the role of this oncogene in SSs and to evaluate the application of inhibitory humanized monoclonal antibodies in the treatment regimens for this malignancy. *HUM PATHOL* 34:639-645. © 2003 Elsevier Inc. All rights reserved.

**Key Words:** HER2/neu, synovial sarcoma, real-time RT-PCR, immunohistochemistry.

**Abbreviations:** FISH, fluorescence in situ hybridization, RT-PCR, reverse transcription-polymerase chain reaction, SS, synovial sarcoma.

Synovial sarcoma (SS) is an aggressive soft tissue tumor that accounts for up to 10% of sarcomas, with a peak incidence in adolescents and young adults. This tumor occurs in 2 major forms, biphasic and monophasic, and it is cytogenetically characterized by the t(X;18)(p11;q11) translocation, found in >95% of cases. Although traditionally considered to be a high-grade neoplasm, recent investigations have suggested that different factors influence prognosis, including morphological and cytogenetic features, treatment strategies, the ploidy status, and the apoptotic index.<sup>1</sup>

The development of new therapeutic advancements, such as the specific targeting of molecular alterations present in human malignancies, has brought to light the

need to identify not only prognostic factors, but also tumor features that are predictive of response to therapy.

One of the most extensively studied molecular targets for therapy is the HER-2/neu proto-oncogene. The HER-2/neu oncogene (also known as c-erbB-2), located on chromosome 17q21, is a member of the tyrosine kinase receptor family and encodes for a 185-kilodalton protein that shows 50% homology with the epidermal growth factor receptor.<sup>2,3</sup> This gene is amplified and/or overexpressed in 20% to 30% of breast carcinomas<sup>4,5</sup> and in various other tumors,<sup>6</sup> and usually is associated with tumor aggressiveness and poor prognosis.<sup>7,8</sup> Several studies have supported the value of HER-2/neu to predict the response to chemotherapy in breast cancer, and the use of recombinant humanized antibodies to HER-2/neu protein (Trastuzumab) in the care of patients with advanced, metastatic breast tumors has been approved.<sup>9</sup>

The role of HER-2/neu activation in soft tissue tumors remains poorly understood, and scarce molecular data backing immunohistochemical studies have been reported. HER-2/neu protein expression was immunohistochemically studied in 204 sarcomas, including 6 SSs, and overexpression was absent in all these malignant mesenchymal neoplasms.<sup>10</sup>

Recently, HER-2/neu alterations have been described in osteosarcoma, with a high incidence of pro-

From the Department of Medicine, Surgery and Dental Sciences, University of Milan, A.O.S. Paolo and IRCCS Ospedale Maggiore, Milan, Italy; Department of Pathology, Orthopedic Institute Gaetano Pini, Milan, Italy; and Interuniversity Center of Cancer Research, Milan, Italy. Accepted for publication March 5, 2003.

Supported by grants from Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR-cofin 1999) and Associazione Italiana per la Ricerca sul Cancro (AIRC).

Address correspondence and reprint requests to Paolo Giovanni Nuciforo, MD, Division of Anatomic Pathology, A.O.S. Paolo, Via A. Di Rudini 8, 20142, Milano, Italy.

© 2003 Elsevier Inc. All rights reserved.

0046-8177/03/3407-0002\$30.00/0

doi:10.1016/S0046-8177(03)00238-7

TABLE 1. Clinicopathologic Features of 13 Patients With SS

No.	Age(y)/G	Site	Subtype	Surgery	Chemo/Radio	Rec	Mets	HER2	FU(mo)
1	68/F	KI	MF	RR	No	Yes	Yes	L	108
2	71/F	KE	BF	WA	No	No	No	L	12
3	27/M	KI	BF	ME	Yes	Yes	Yes	L	24
4	39/M	T	MF	WE	Yes	No	No	H	48
5	15/M	N	MF,PD	WE	Yes	No	No	H	48
6	41/F	F	MF,PD	WA	Yes	No	No	H	48
7	57/F	KI	MF	WA	No	N/A	N/A	L	N/A
8	48/M	F	BF,PD	ME	Yes	Yes	No	L	36
9	29/F	T	PD	WE	Yes	Yes	Yes	L	36
10	27/M	KE	BF	WE	No	No	No	H	24
11	62/F	A	MF	WE	No	Yes	No	L	24
12	62/F	F	MF	WA	No	No	No	L	2
13	34/F	L	MF	RR	No	No	No	L	2

Abbreviations: Age, age at diagnosis; G, gender; M, male; F, female; Site, anatomic location; KI, knee, intra-articular; KE, knee, extra-articular; T, thigh; N, neck; F, foot; A, arm; L, leg; Subtype, histological subtype; MF, monophasic fibrous; BF, biphasic; MF-PD, monophasic fibrous with poorly differentiated areas; BF-PD, biphasic with poorly differentiated areas; PD, poorly differentiated; Surgery, primary surgical therapy; RR, radical en bloc resection; WA, wide through-bone amputation; ME, marginal en bloc excision; WE, wide en bloc excision; Chemo/Radio, adjuvant postoperative chemotherapy and/or radiotherapy; Rec, recurrence; Mets, presence of metastasis; HER2, HER2 mRNA expression; L, low expression; H, high expression; FU, follow-up status; N/A, not available.

tein expression, ranging from 42% to 61%.<sup>11-15</sup> Indeed, despite limited information on Her2/neu in this type of malignancy, based mostly on immunohistochemical findings, 2 clinical trials of Trastuzumab have been initiated for recurrent and metastatic osteosarcoma patients ([http://www.cancer.gov/clinical\\_trials](http://www.cancer.gov/clinical_trials): MSKCC-99097/NCI-T98-0083 and COG-AOST0121).

Therefore, we evaluated the mRNA expression and the gene product of HER-2/neu in 13 SS patients using real-time reverse transcription-polymerase chain reaction (RT-PCR) and immunohistochemistry. To the best of our knowledge, this is the first report documenting increased levels of HER-2/neu mRNA and protein in SS.

## MATERIALS AND METHODS

### Patient Population

Thirteen patients with primary SSs, obtained from the files of the Department of Pathology, Gaetano Pini Orthopedic Institute, were included in this study. Cases were chosen based on the availability of frozen primary tumor. Patient age ranged from 15 to 71 years (mean, 44.6 years). Anatomic sites included the knee (5 patients; 3 intra-articular), thigh (2 patients), foot (3 patients), arm (1 patient), neck (1 patient), and lower leg (1 patient). The histological subtypes were 3 biphasic (BF), 6 monophasic fibrous (MF), 2 monophasic fibrous with poorly differentiated areas (MF-PD), 1 biphasic with poorly differentiated areas (BF-PD), and 1 poorly differentiated (PD). Clinical staging was IIB for all the patients. Local surgical excision was performed in 9 patients; amputation, in 4 patients. Recurrence was observed in 5 cases with subsequent amputation (3 patients) and local excision (2 patients). Metastasis occurred to lung (2 patients) and inguinal lymph nodes (1 patient). Five patients received adjuvant postoperative chemotherapy, and 1 patient (case 4) was treated with chemotherapy associated with radiotherapy. Follow-up ranged from 2 to 108 months (mean, 34.3 months). Clinicopathologic data are summarized in Table 1.

### Pathologic Studies and Frozen Tissue Selection

In all cases, the primary tumor was available for study. Surgically resected tumor tissues were partly snap-frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  for RNA extraction, and partly fixed in buffered formalin and embedded in paraffin blocks. Hematoxylin and eosin-stained sections were re-evaluated and graded according to FNCLCC grading system.<sup>16</sup> Frozen tissue blocks were handled as follows: 4- $\mu$ -thick frozen sections were cut and stained with hematoxylin and eosin to determine the percentage of tumor cells present in the specimen. We used tissue blocks with tumor cells comprising more than 80% of the specimen. About 10 20- $\mu$ -thick sections were collected into Eppendorf tubes. Another 4- $\mu$ -thick frozen section was cut after the serial sections and examined by light microscopy to guarantee the percentage of tumor cells collected. Trizol (Life Technologies; Gibco BRL, Gaithersburg, MD) was used for RNA extraction, according to the manufacturer's protocol. RNA was quantified spectrophotometrically.

### cDNA Synthesis

Total RNA (200 ng) was reverse-transcribed in a total volume of 50  $\mu\text{L}$  containing 1  $\times$  TaqMan buffer, 5.5 mmol  $\text{MgCl}_2$ , 1 mmol deoxynucleotides, 2.5  $\mu\text{mol}$  random hexamers, 20 U RNase inhibitor, and 62.5 U MuL reverse transcriptase. The samples were incubated at  $25^{\circ}\text{C}$  for 10 minutes,  $48^{\circ}\text{C}$  for 30 minutes and  $95^{\circ}\text{C}$  for 5 minutes.

### PCR Amplification

Amplification reactions were performed with the Universal TaqMan 2  $\times$  PCR mastermix in a volume of 25  $\mu\text{L}$  containing 300 nmol of each primer, 100 nmol of probe, and 5  $\mu\text{L}$  of cDNA. Both  $\beta$ -actin and HER2/neu amplification were done in duplicate for each sample.

The thermal cycling conditions included 2 minutes at  $50^{\circ}\text{C}$  and 10 minutes at  $95^{\circ}\text{C}$ , followed by 40 cycles of  $95^{\circ}\text{C}$  for 15 seconds and  $60^{\circ}\text{C}$  for 1 minute. All reagents used for



RT-PCR were purchased from Applied Biosystems (Foster City, CA).

#### Primers and Probes

Primers and probes for  $\beta$ -actin and HER2/neu mRNA were chosen using the computer program Primer Express (Applied Biosystems). Sequences of the forward primer for HER2/neu mRNA (GenBank accession number X03363) were 5'-TCC TGT GTG CAC CTG CAT CAC-3' and the reverse primer 5'-CCA AAG ACC ACC CCC AAG A-3'; the sequence of the TaqMan probe was 5'(FAM)-ACC AGA ATG CCA ACC ACC GCA GA-(TAMRA)-3'. Sequences of the forward primer for  $\beta$ -actin mRNA (GenBank accession number X00351) were 5-TCC TTC CTG CCC ATG GAG-3' and the reverse primer 5'-ACC ACG ACC AAT GAT CTT GAT CTT-3'; the sequence of the TaqMan probe was 5'(FAM)-CCT GTG CCA TCC ACC AAA CTA CCT TG-(TAMRA)-3'. Probes were purchased from Applied Biosystems.

#### Real-Time RT-PCR

To measure HER2/neu expression in these tumors we used a real-time quantitative RT-PCR based on TaqMan methodology, as previously described,<sup>17</sup> with minor modifications. Briefly, this technique allows, by means of fluorescence emission, to find the cycling point when PCR product is detectable (Ct value or threshold cycle). As previously reported, the Ct value correlates to the starting quantity of the target mRNA.<sup>18</sup> To normalize the amount of total RNA present in each reaction, we amplified the housekeeping gene  $\beta$ -actin, which is assumed to be constant in both normal samples and tumor tissues.

Our results are expressed as relative levels of HER2/neu mRNA, referred to a sample, called a "calibrator," chosen to represent 1X expression of this gene. The calibrator was a breast cancer cellular line (MCF-7)<sup>19</sup> that was analyzed on every assay plate with the unknown samples. All of the analyzed tumors expressed n-fold HER2/neu mRNA relative to the calibrator.

The amount of target, normalized to an endogenous reference ( $\beta$ -actin) and relative to the calibrator, was defined by the  $\Delta\Delta Ct$  method as described by Livak K (Sequence Detector User Bulletin 2; Applied Biosystems). Specifically, the formula is applied as follows:

$$\text{target amount} = 2^{-\Delta\Delta Ct}$$

where  $\Delta\Delta Ct = [Ct (\text{HER2/neu sample}) - Ct (\beta\text{-actin sample})] - [Ct (\text{HER2/neu calibrator}) - Ct (\beta\text{-actin calibrator})]$ .

#### Immunohistochemistry

Formalin-fixed, paraffin-embedded tissue sections were deparaffinized, rehydrated, and exposed to the primary antibody using the EnVision+ system (Dako, Carpinteria, CA). Primary anti-HER2/neu antibody (rabbit polyclonal antibody, catalog number A0485; Dako) was applied in a dilution of 1:2000 for 60 minutes at room temperature. Before exposure to the primary antibody, sections were microwave-pre-treated in EDTA, pH 8.0, to retrieve antigenicity, and incubated with endogenous peroxidase-blocking solution for 10 minutes at room temperature. Positive control, constituted by a breast carcinoma showing more than 80% positive staining for HER2/neu, as well as negative control, in which the primary antibody was omitted, were stained in parallel.

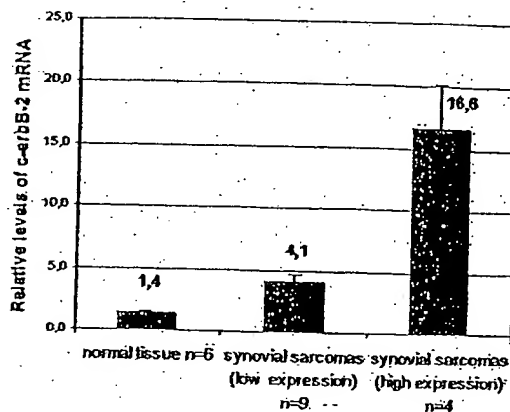


FIGURE 1. Distribution of HER2/neu mRNA levels in normal muscle tissues, and in low- and high-expression sarcomas. Data are expressed as mean and standard error of the mean for each group.

All cases were examined for both cytoplasmic and membrane immunoreactivity. Cytoplasmic staining was evaluated on a semiquantitative scale, according to Kilpatrick et al with minor modifications,<sup>20</sup> and reported as 0 (no staining or staining in <10% of cells), 1+ (weak staining in >10% of cells), 2+ (moderate staining in >10% of cells), or 3+ (strong staining in >10% of cells). The presence of a membranous pattern of staining was recorded separately and scored as absent (no staining or weak staining in <10% of cells) or present (complete and/or incomplete staining in >10% of cells). Tumors with a cytoplasmic score of 3+ were considered to have high HER2/neu protein expression.

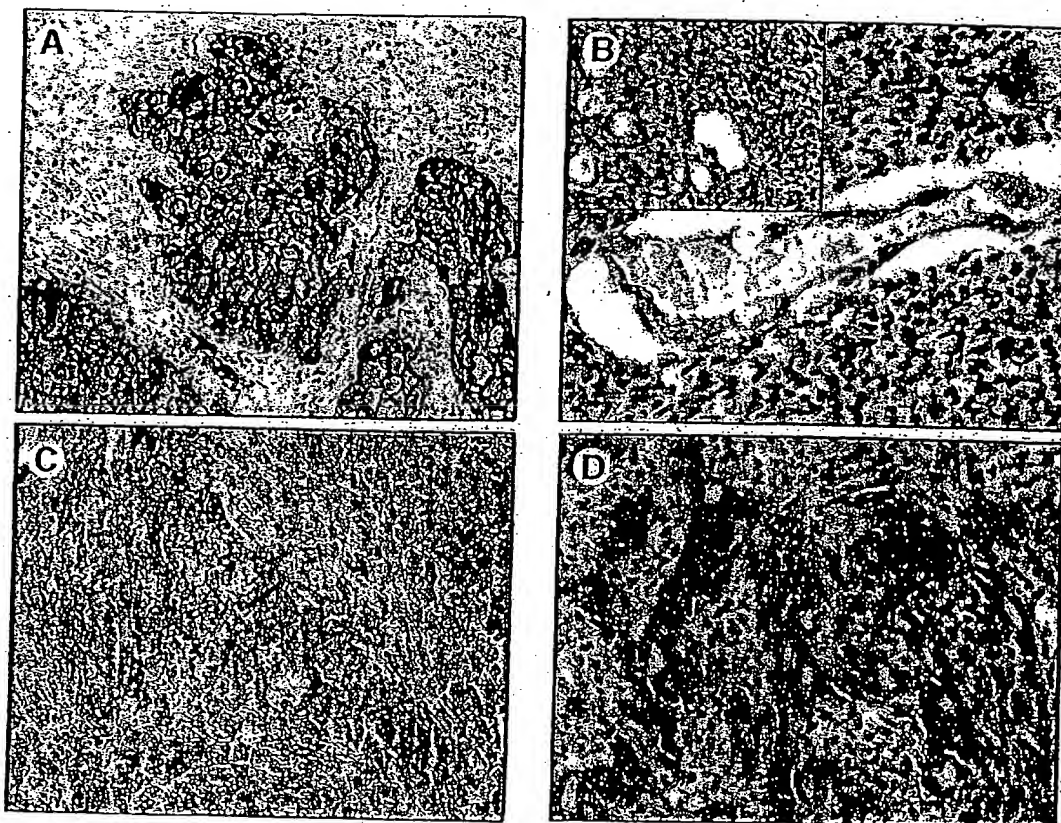
#### Statistical Analysis

Statistical differences were calculated by Fisher's exact test. The t-test method was used to evaluate the differences between groups. Differences were considered statistically significant when  $P$  was <0.05.

#### RESULTS

##### HER2/neu mRNA Evolution

All of the tissues analyzed contained detectable levels of HER2/neu mRNA. Six normal tissue samples (skeletal muscle) were used to establish basal level of HER2/neu mRNA. All the normal samples expressed very low levels of HER2/neu mRNA, ranging from 0.9 to 1.9 n (mean, 1.4 n). Among the 13 tumor samples, HER2/neu levels varied greatly, ranging from 2.1 to 24 n. Setting a cutoff level at 7.9 n (a value that represents the mean value of expression distribution of the SSs), 9 cases (69%) had low HER2/neu expression and 4 cases (31%) had high HER2/neu expression (Fig 1; Table 1). The difference between the 2 groups (low and high HER2/neu tumors) was statistically significant ( $P = 0.0001$ ).



**FIGURE 2.** Immunohistochemical localization of HER2/neu in SS. (A) Positive control (breast cancer) showing typical strong membrane pattern of positivity. (B) Case 2, a biphasic SS, displaying very focal membrane staining limited to a gland. Inset: focal weak cytoplasmic positivity in the epithelial component of the same case. (C) Case 10, the epithelial component of this biphasic SS, displaying strong membrane pattern of staining. (D) Case 4, an epithelioid area in a monophasic fibrous SS, showing strong cytoplasmic positivity.

#### HER2/neu Protein Expression

Staining with HER2/neu antibody revealed a variable cytoplasmic and membrane staining pattern. Three tumors (23%) showed strong staining involving both the cell membrane and the cytoplasm (cases 4, 5, and 10); weak to moderate, exclusively cytoplasmic staining was observed in 7 cases (cases 1, 3, 6, 7, 8, 9, and 11). No staining was detected in 2 tumors (cases 12 and 13). In 1 case (case 2), a cluster of glandular structures representing <5% of the tumor showed weak cytoplasmic and very focal membrane staining, the latter limited to a single gland. The epithelial/epithelioid components exhibited stronger cytoplasmic staining compared with the spindle-cell component of the tumors. Membrane staining was predominantly incomplete and limited to the epithelial/epithelioid areas. All 3 cases with high immunohistochemical expression of HER2/neu (cases 4, 5, and 10) were grade III sarcomas, including 1 MF, 1 BF, and 1 MF-PD SSs.

Examples of HER2/neu cytoplasmic and membrane staining are depicted in Figure 2.

#### Correlation of Molecular and Immunohistochemical Results

A strong, statistically significant association was present between protein expression, for both membrane and cytoplasmic staining, and HER2/neu mRNA levels ( $P = 0.01$ ), although 1 case (case 6) displayed discordant results. Interestingly, this neoplasm showed high HER2/neu mRNA levels, whereas only weak staining, limited to the cytoplasm of a minority of tumor cells, was detected by immunohistochemical analysis.

#### HER2/neu Expression and Clinicopathologic Parameters

Both HER2/neu protein expression and mRNA levels were evaluated to establish the relationships to

**TABLE 2.** Correlation Between Clinicopathologic Features and HER2/neu Expression as Detected by IHC and RT-PCR

Variable	HER2/neu					
	IHC			PCR		
	L	H	P value	L	H	P value
Age (years)						
<40	3	3		3	3	
>40	7	0	NS	6	1	NS
Sex						
Female	8	0		7	1	
Male	2	3	0.03	2	3	NS
Tumor size (cm)						
<5	3	2		3	2	
>5	7	1	NS	6	2	NS
Histological grade						
II	3	0		3	0	
III	7	3	NS	6	4	NS
Histological type						
MF	5	1		5	1	
BF	2	1		2	1	
PD	3	1	NS	2	2	NS
Chemo/Radiotherapy*						
Yes	4	2		3	3	
No	5	1	NS	5	1	NS
Recurrence†						
Yes	5	0		5	0	
No	2	3	NS	1	4	0.02
Metastasis‡						
Yes	3	0		3	0	
No	4	3	NS	3	4	NS

Abbreviations: L, low expression; H, high expression; NS, not significant; MF, monophasic fibrous; BF, biphasic; PD, poorly differentiated (including MF and BF with poorly differentiated areas).

\*Information not available for case 7.

†Cases 7, 12, and 13 were excluded from the analysis.

clinicopathologic features, including local recurrence and metastatic disease. Two cases (cases 12 and 13) with follow-up less than 12 months and 1 case (case 7) for which clinical information was not available were excluded from the analysis of recurrences and metastatic behavior.

No correlation was observed between HER2/neu mRNA expression and age, sex, tumor size, tumor grade, histotype, and metastasis. A correlation between sex of the patients and HER2/neu protein expression was found. In fact, none of the female patients showed high HER2/neu protein expression ( $P = 0.03$ ). Patients with high Her2/neu mRNA levels had a lower risk of recurrence than those with low Her2/neu mRNA levels ( $P = 0.02$ ). None of the cases with high HER2/neu mRNA levels developed metastatic foci, although the small number of observations precluded reaching statistical significance ( $P = 0.1$ ). Results are detailed in Table 2.

## DISCUSSION

The present work provides the first combined molecular by real-time RT-PCR and immunohistochemical evidence that HER2/neu overexpression occurs in SSs.

Our results indicate that this parameter may provide prognostic information and suggest that a specific therapy with humanized monoclonal antibodies against HER2/neu may be considered in a significant number of SSs.

The HER2/neu oncogene has been extensively investigated as a prognostic factor and more recently as a predictor of response to therapy. It has been demonstrated in breast cancer, where HER2/neu overexpression is usually associated with gene amplification,<sup>21</sup> and in other epithelial tumors, including ovarian, gastric, lung, and urinary bladder carcinomas.

HER2/neu amplification/overexpression appears to be an early event in oncogenic transformation by interacting with other members of the HER family.<sup>3</sup> In breast cancer, it is involved in cell cycle and apoptotic pathways through the antiapoptotic effects mediated by p53 and p21 deregulation.<sup>22,23</sup>

Whether HER2/neu overexpression plays an important role in mesenchymal neoplasms remains controversial. An immunohistochemical study of sarcomas, using a monoclonal antibody, reported no evidence of immunoreactivity for HER2/neu in 6 SSs as well as in other 197 mesenchymal tumors, with cytoplasmic reactivity observed only in 1 case of peripheral neuroepithelioma.<sup>10</sup> A recent investigation reported gene expression profiles of 41 soft tissue tumors with cDNA microarray analysis. Among these sarcomas, 6 monophasic SSs were characterized by a unique expression pattern of a cluster of 104 genes, including the epidermal growth factor receptor, which shows 50% homology with the HER2/neu gene.<sup>24</sup> These data also suggest that the erb-B receptor family plays a significant role in SS. It has been demonstrated that a variable number of osteosarcomas overexpress HER2/neu.<sup>11-15</sup> However, more recent studies<sup>20,25,26</sup> were unable to detect any HER2/neu gene amplification and/or overexpression using fluorescence in situ hybridization (FISH), RT-PCR, and immunohistochemistry.

Differences in the techniques used may play an important role and explain (at least in part) these discrepancies. HER2/neu alterations can be evaluated using different techniques including immunohistochemistry, FISH, Southern hybridization, Northern blot, and competitive, differential, or real-time PCR.<sup>27</sup> Immunohistochemistry is the most common method for detection of HER2/neu overexpression, but it is significantly affected by the sensitivity and specificity of the antibodies used, the type of tissue (frozen versus formalin-fixed), and the various interpretative criteria and scoring systems used to evaluate cases. Indeed, most studies of HER2/neu expression in osteosarcoma used immunohistochemical techniques, with different monoclonal or polyclonal antibodies. The discrepancy in results may stem from the use of different antibodies, as well as a lack of standardized evaluation.

For these reasons, to evaluate HER2/neu immunoreactivity in our study, we used a polyclonal antibody (Dako, Carpinteria, CA), arguably the most diffuse and thoroughly tested antibody for HER2/neu assessment. Furthermore, we investigated HER2/neu mRNA ex-

pression with real-time RT-PCR, because it has been demonstrated that mRNA levels correlates tightly with protein expression.<sup>28</sup> At present, real-time RT-PCR probably represents the most powerful tool for quantitative analysis, because it allows better internal control and reduction of sample contamination, and provides more objective results.<sup>18</sup>

We analyzed HER2/neu gene expression at the mRNA and protein level in 13 cases of SS. HER2/neu expression was found in all of the cases investigated, and mRNA content in the tumors varied from 2.1 to 24 n.

The variability of mRNA levels in SSs is reflected on heterogeneity of protein expression pattern as detected by immunohistochemistry. We found that HER2/neu immunoreactivity correlates strongly with mRNA levels. A convincing cytoplasmic immunoreactivity was documented with the polyclonal antibody in 10 of 13 sarcomas. Distinct membranous staining was observed in 3 cases, although it was never comparable to the positive breast control. It was predominantly incomplete and identified in the epithelial/epithelioid component of SS. Interestingly, all of the cases with strong cytoplasmic staining also exhibited a membrane-staining pattern.

In breast cancers, a membranous pattern of staining is thought to be specific for HER2/neu protein expression and correlated with gene amplification, whereas cytoplasmic staining is usually considered non-specific.<sup>29</sup> However, cytoplasmic positivity for HER2/neu has been reported to be prognostically significant in other tumor types, including bladder, colon, pancreas, thyroid, and nasopharyngeal carcinomas, and even in breast cancer.<sup>30,35</sup>

Patients with high HER2/neu mRNA expression had a significantly lower risk of recurrence. Similarly, all of the cases with high HER2/neu expression did not metastasize, although this correlation did not reach statistical significance due to the small number of cases studied. These data suggest that HER2/neu plays a role in the biology of SS and that HER2/neu overexpression may be linked to a less-aggressive clinical behavior. Indeed, unlike many cancers where HER2/neu overexpression has been shown to correlate with poor prognosis, HER2/neu levels are linked to a more favorable clinical course in other malignant tumors, such as thyroid carcinoma and osteosarcoma.<sup>15,34</sup>

The molecular mechanisms responsible for the action of HER-2/neu in SSs are unknown. A possible interaction between HER-2/neu and the other members of HER family could be important in tumorigenesis. Derangements of other oncogenes, tumor suppressors, and apoptosis regulators have been described in SSs. For instance, many SSs have been shown to be diffusely positive for bcl-2 family proteins (bcl-2, bax, bcl-x, and bak). These members of the bcl-2 family are involved in the regulation of apoptosis in SS.<sup>36</sup> This raises the hypothesis that complex alterations in apoptosis-controlling mechanisms are present in these neoplasms, with HER-2/neu interacting with Bcl-2 family

members. Further studies are needed to clarify the mechanisms of apoptosis in SS.

Depending on the size and location, the therapy of choice for SS is radical local excision or amputation. Whenever radical surgery cannot be performed, radiotherapy in concert with local excision is suggested in an attempt to avoid the need for amputation. Only recently has a study evaluated the possible role of chemotherapy in the treatment of SS.<sup>37</sup>

The role of molecular markers in predicting treatment responsiveness is currently the focus of extensive investigation. Breast cancer patients with high HER2/neu expression appeared to benefit from high-dose CAF (cyclophosphamide, adriamycin, and 5-fluorouracil) therapy.<sup>38</sup> In our study, 3 of 4 patients with high HER2/neu expression received adjuvant chemotherapy with ifosfamide; these patients had a favorable clinical outcome. These data raise the possibility that HER2/neu may have value in predicting which patients are likely to respond to specific adjuvant chemotherapy regimens. Whether the favorable significance of HER2/neu expression depends on predicting clinical recurrence, response to chemotherapy, or both remains to be fully elucidated in SS patients.

To the best of our knowledge, this is the first report that shows expression of HER2/neu in primary SS by real-time RT-PCR. Elevated levels of HER2/neu mRNA and protein are found in a significant group of SS patients, and these levels appear to correlate with features of good prognosis. Furthermore, our results suggest that this mechanism of disease in SS may be the target of specific inhibitory therapies based on humanized monoclonal antibodies. Considering the small number of patients examined, further investigation is needed to confirm these preliminary findings.

## REFERENCES

1. Weiss SW, Goldblum JR: Synovial sarcoma, in Weiss SW, Goldblum JR (eds): *Enzinger and Weiss's Soft Tissue Tumors*, 4th ed. St Louis, MO, Mosby-Year Book, 2001, pp 1483-1509
2. Schechter AL, Stern DF, Vaidyanathan L, et al: The neu oncogene: An erb-B-related gene encoding an 185,000 Mr tumor antigen. *Nature* 312:513-516, 1984
3. Bargmann CI, Hung MC, Weinberg RA: The neu oncogene encodes an epidermal growth factor receptor-related protein. *Nature* 319:226-230, 1986
4. Ali IU, Campbell C, Lidereau R, et al: Lack of evidence for the prognostic significance of c-erbB-2 amplification in human breast carcinoma. *Oncogene Res* 3:139-146, 1988
5. Berger MS, Locher GW, Saurer S, et al: Correlation of c-erbB-2 gene amplification and protein expression in human breast carcinoma with nodal status and nuclear grading. *Cancer Res* 48:1238-1243, 1988
6. Scholl S, Beuzebat P, Pouillart P: Targeting HER2 in other tumor types. *Ann Oncol* 12(Suppl1):S81-S87, 2001
7. Slamon DJ, Clark GM, Wong SG, et al: Human breast cancer: Correlation to relapse and survival with amplification of the HER-2/neu oncogene. *Science* 235:177-182, 1987
8. Wright C, Angus B, Nicholson S, et al: Expression of c-erbB2 oncoprotein: A prognostic indicator in human breast cancer. *Cancer Res* 49:2087-2090, 1989
9. Slamon D, Leyland-Jones B, Shak S, et al: Use of chemotherapy plus a monoclonal antibody against Her2/neu for metastatic

- breast cancer that overexpress Her2/neu. *N Engl J Med* 344:783-792, 2001
10. George E, Nichols GA, Swanson PE, et al: Overexpression of the c-erbB-2 oncogene in sarcomas and small round cell tumors of childhood. An immunohistochemical investigation. *Arch Pathol Lab Med* 116:1033-1035, 1992
11. Onda M, Matsuda S, Higaki S, et al: ErbB-2 expression is correlated with poor prognosis for patients with osteosarcoma. *Cancer* 77:71-78, 1996
12. Gorlick R, Huvois AG, Heller G, et al: Expression of HER2/erbB-2 correlates with survival in osteosarcoma. *J Clin Oncol* 17:2781-2788, 1999
13. Akatsuka T, Wada T, Kokai Y, et al: Loss of ErbB2 expression in pulmonary metastatic lesions in osteosarcoma. *Oncology* 60:361-366, 2001
14. Morris CD, Gorlick R, Huvois G, et al: Human epidermal growth factor receptor 2 as a prognostic indicator in osteogenic sarcoma. *Clin Orthop* 382:59-65, 2001
15. Akatsuka T, Wada T, Kokai Y, et al: ErbB2 expression is correlated with increased survival of patients with osteosarcoma. *Cancer* 94:1397-1404, 2002
16. Coindre JM, Trojani M, Contesso G, et al: Reproducibility of a histopathologic grading system for adult soft tissue sarcoma. *Cancer* 58:306-309, 1986
17. Marchetti A, Pellegrini C, Buttitta F, et al: Prediction of survival in stage I lung carcinoma patients by telomerase function evaluation. *Lab Invest* 82:729-736, 2002
18. Heid CA, Stevens J, Livak KJ, et al: Real time quantitative PCR. *Genome Res* 6:986-994, 1996
19. Kraus MH, Popescu NC, Amsbaugh SC, et al: Overexpression of the EGF receptor-related proto-oncogene erbB-2 in human mammary tumor cell lines by different molecular mechanisms. *EMBO J* 6:605-610, 1987
20. Kilpatrick SE, Geisinger KR, King TS, et al: Clinicopathologic analysis of HER-2/neu immunoreactivity among various histologic subtypes and grades of osteosarcoma. *Mod Pathol* 14:1277-1283, 2001
21. Slamon DJ, Godolphin W, Jones LA, et al: Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science* 244:707-712, 1989
22. Harari D, Yarden Y: Molecular mechanisms underlying ErbB2/HER2 action in breast cancer. *Oncogene* 19:6102-6114, 2000
23. Zhou BP, Liao Y, Xia W, et al: HER-2/neu induces p53 ubiquitination via Akt-mediated MDM2 phosphorylation. *Nat Cell Biol* 3:973-982, 2001
24. Nielsen OT, West RB, Linn SC, et al: Molecular characterization of soft tissue tumours: A gene expression study. *Lancet* 359: 1301-1307, 2002
25. Thomas DG, Giordano TJ, Sanders D, et al: Absence of HER2/neu gene expression in osteosarcoma and skeletal Ewing's sarcoma. *Clin Cancer Res* 8:788-793, 2002
26. Maitra A, Wanzer D, Weinberg AG, et al: Amplification of the Her-2/neu oncogene is uncommon in pediatric osteosarcomas. *Cancer* 92:677-683, 2001
27. Hanna W, Kahn HJ, Trudeau M: Evaluation of HER-2/neu (erbB-2) status in breast cancer: From bench to bedside. *Mod Pathol* 12:827-834, 1999
28. Naber SP, Tsutsumi Y, Yin S, et al: Strategies for the analysis of oncogene overexpression. Studies of the neu oncogene in breast carcinoma. *Am J Clin Pathol* 94:125-136, 1990
29. Teto B, Brisson J: Prognostic significance of HER-2/neu oncoprotein expression in node-positive breast cancer. The influence of the pattern of immunostaining and adjuvant therapy. *Cancer* 73:2359-2365, 1994
30. Coombs LM, Figott DA, Sweeney E, et al: Amplification and overexpression of c-erbB-2 in transitional cell carcinoma of the urinary bladder. *Br J Cancer* 63:601-608, 1991
31. Kay EW, Mulcahy H, Walsh CB, et al: Cytoplasmic c-erbB-2 protein expression correlates with survival in Dukes' B colorectal carcinoma. *Histopathology* 25:455-461, 1994
32. Roychowdhury DF, Tseng A Jr, Fu KK, et al: New prognostic factors in nasopharyngeal carcinoma. Tumor angiogenesis and c-erbB-2 expression. *Cancer* 77:1419-1426, 1996
33. Hall PA, Hughes CM, Staddon SL, et al: The c-erb B-2 proto-oncogene in human pancreatic cancer. *J Pathol* 161:195-200, 1990
34. Sugg SL, Ezzat S, Zheng L, et al: Cytoplasmic staining of erbB-2 but not mRNA levels correlates with differentiation in human thyroid neoplasia. *Clin Endocrinol (Oxf)* 49:629-637, 1998
35. Keshgegian AA, Cnaan A: erbB-2 oncoprotein expression in breast carcinoma. Poor prognosis associated with high degree of cytoplasmic positivity using CB-11 antibody. *Am J Clin Pathol* 108: 456-463, 1997
36. Kawauchi S, Fukuda T, Oda Y, et al: Prognostic significance of apoptosis in synovial sarcoma: Correlation with clinicopathologic parameters, cell proliferative activity, and expression of apoptosis-related proteins. *Mod Pathol* 13:755-765, 2000
37. Rosen G, Forscher C, Lowenbraun S, et al: Synovial sarcoma: Uniform response of metastasis to high dose ifosfamide. *Cancer* 73:2506-2511, 1994
38. Muss HB, Thor AD, Berry DA, et al: c-erbB-2 expression and response to adjuvant therapy in women with node-positive early breast cancer. *N Engl J Med* 330:1260-1266, 1994

401: Biochem Biophys Res Commun. 1995 Sep 25;214(3):1009-14.

[Related Articles, Links](#)

ELSEVIER  
JOURNAL

**Differential expression of heat shock protein 70 in well healing and chronic human wound tissue.**

Oberringer M, Baum HP, Jung V, Welter C, Frank J, Kuhlmann M, Mutschler W, Hanselmann RG.

Department of Traumatology, University of Saarland, Germany.

Heat shock protein 70 (hsp 70) is an important member of the heat shock protein family, which is induced by different forms of stress. We attempted to find out if hsp 70 is also involved in wound healing, which likewise resembles a stress situation for cells too. Therefore we collected tissue samples from well healing and chronic human wound tissue. We used Northern- and Western-blot analysis to study the expression of hsp 70. At the protein level we found a strong correlation between well healing wounds and high expression of hsp 70, whereas chronic wounds showed no or weak expression. Interestingly hsp 70 mRNA did not show this significant correlation, displaying a variant expression pattern in the same kind of wound tissue, possibly due to unknown posttranscriptional regulating step, which has to be investigated in further studies. To localize hsp 70 mRNA and protein was used insitu hybridization and immunohistochemistry. Both displayed an overexpression in endothelial cells of capillary vessels.



**Genome-wide study of gene copy numbers, transcripts, and protein levels in pairs of non-invasive and invasive human transitional cell carcinomas.**

Orntoft TF, Thykjaer T, Waldman FM, Wolf H, Celis JE.

Department of Clinical Biochemistry, Molecular Diagnostic Laboratory, Aarhus University Hospital, Skejby, DK-8200 Aarhus N, Denmark. orntoft@kba.sks.au.dk

Gain and loss of chromosomal material is characteristic of bladder cancer, as well as malignant transformation in general. The consequences of these changes at both the transcription and translation levels is at present unknown partly because of technical limitations. Here we have attempted to address this question in pairs of non-invasive and invasive human bladder tumors using a combination of technology that included comparative genomic hybridization, high density oligonucleotide array-based monitoring of transcript levels (5600 genes), and high resolution two-dimensional gel electrophoresis. The results showed that there is a gene dosage effect that in some cases superimposes on other regulatory mechanisms. This effect depended ( $p < 0.015$ ) on the magnitude of the comparative genomic hybridization change. In general (18 of 23 cases), chromosomal areas with more than 2-fold gain of DNA showed a corresponding increase in mRNA transcripts. Areas with loss of DNA, on the other hand, showed either reduced or unaltered transcript levels. Because most proteins resolved by two-dimensional gels are unknown it was only possible to compare mRNA and protein alterations in relatively few cases of well focused abundant proteins. With few exceptions we found a good correlation ( $p < 0.005$ ) between transcript alterations and protein levels. The implications, as well as limitations, of the approach are discussed.

PMID: 12096139 [PubMed - indexed for MEDLINE]

# Genome-wide Study of Gene Copy Numbers, Transcripts, and Protein Levels in Pairs of Non-invasive and Invasive Human Transitional Cell Carcinomas\*

Torben F. Ørntoft†§, Thomas Thykjaer||, Frederic M. Waldman||, Hans Wolf\*\*, and Julio E. Celis††

Gain and loss of chromosomal material is characteristic of bladder cancer, as well as malignant transformation in general. The consequences of these changes at both the transcription and translation levels is at present unknown partly because of technical limitations. Here we have attempted to address this question in pairs of non-invasive and invasive human bladder tumors using a combination of technology that included comparative genomic hybridization, high density oligonucleotide array-based monitoring of transcript levels (5600 genes), and high resolution two-dimensional gel electrophoresis. The results showed that there is a gene dosage effect that in some cases superimposes on other regulatory mechanisms. This effect depended ( $p < 0.015$ ) on the magnitude of the comparative genomic hybridization change. In general (18 of 23 cases), chromosomal areas with more than 2-fold gain of DNA showed a corresponding increase in mRNA transcripts. Areas with loss of DNA, on the other hand, showed either reduced or unaltered transcript levels. Because most proteins resolved by two-dimensional gels are unknown it was only possible to compare mRNA and protein alterations in relatively few cases of well focused abundant proteins. With few exceptions we found a good correlation ( $p < 0.005$ ) between transcript alterations and protein levels. The implications, as well as limitations, of the approach are discussed. *Molecular & Cellular Proteomics* 1:37–45, 2002.

Aneuploidy is a common feature of most human cancers (1), but little is known about the genome-wide effect of this

phenomenon at both the transcription and translation levels. High throughput array studies of the breast cancer cell line BT474 has suggested that there is a correlation between DNA copy numbers and gene expression in highly amplified areas (2), and studies of individual genes in solid tumors have revealed a good correlation between gene dose and mRNA or protein levels in the case of c-erb-B2, cyclin D1, *ems1*, and N-myc (3–5). However, a high cyclin D1 protein expression has been observed without simultaneous amplification (4), and a low level of c-myc copy number increase was observed without concomitant c-myc protein overexpression (6).

In human bladder tumors, karyotyping, fluorescent *in situ* hybridization, and comparative genomic hybridization (CGH)<sup>1</sup> have revealed chromosomal aberrations that seem to be characteristic of certain stages of disease progression. In the case of non-invasive pTa transitional cell carcinomas (TCCs), this includes loss of chromosome 9 or parts of it, as well as loss of Y in males. In minimally invasive pT1 TCCs, the following alterations have been reported: 2q–, 11p–, 1q+, 11q13+, 17q+, and 20q+ (7–12). It has been suggested that these regions harbor tumor suppressor genes and oncogenes; however, the large chromosomal areas involved often contain many genes, making meaningful predictions of the functional consequences of losses and gains very difficult.

In this investigation we have combined genome-wide technology for detecting genomic gains and losses (CGH) with gene expression profiling techniques (microarrays and proteomics) to determine the effect of gene copy number on transcript and protein levels in pairs of non-invasive and invasive human bladder TCCs.

## EXPERIMENTAL PROCEDURES

**Material**—Bladder tumor biopsies were sampled after informed consent was obtained and after removal of tissue for routine pathology examination. By light microscopy tumors 335 and 532 were staged by an experienced pathologist as pTa (superficial papillary).

<sup>1</sup> The abbreviations used are: CGH, comparative genomic hybridization; TCC, transitional cell carcinoma; LOH, loss of heterozygosity; PA-FABP, psoriasis-associated fatty acid-binding protein; 2D, two-dimensional.

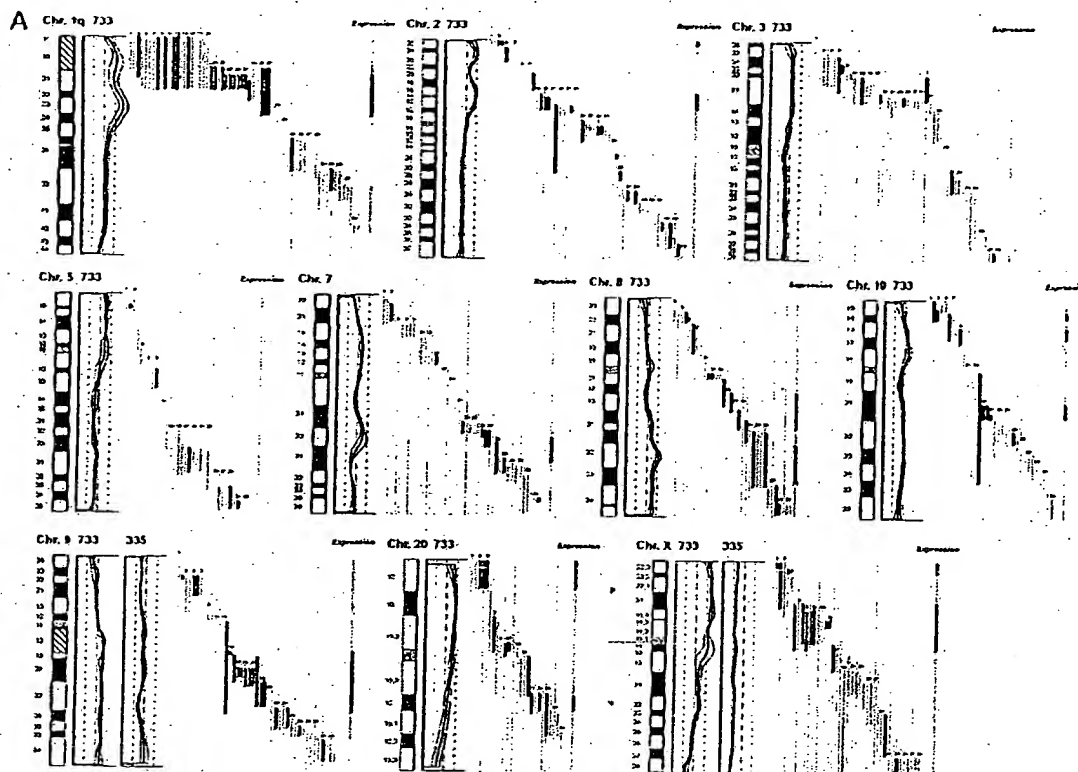
From the †Department of Clinical Biochemistry, Molecular Diagnostic Laboratory and \*\*Department of Urology, Aarhus University Hospital, Skejby, DK-8200 Aarhus N, Denmark, ‡AROS Applied Biotechnology ApS, Gustav Wiedsvej 10, DK-8000 Aarhus C, Denmark, §UCSF Cancer Center and Department of Laboratory Medicine, University of California, San Francisco, CA 94143-0808, and ††Institute of Medical Biochemistry and Danish Centre for Human Genome Research, Ole Worms Allé 170, Aarhus University, DK-8000 Aarhus C, Denmark

Received, September 26, 2001, and in revised form, November 7, 2001

\* Published, MCP Papers in Press, November 13, 2001, DOI 10.1074/mcp.M100019-MCP200



# Gene Copy Numbers, Transcripts, and Protein Levels



**Fig. 1.** DNA copy number and mRNA expression level. Shown from left to right are chromosome (Chr.), CGH profiles, gene location and expression level of specific genes, and overall expression level along the chromosome. A, expression of mRNA in invasive tumor 733 as compared with the non-invasive counterpart tumor 335. B, expression of mRNA in invasive tumor 827 compared with the non-invasive counterpart tumor 532. The average fluorescent signal ratio between tumor DNA and normal DNA is shown along the length of the chromosome (left). The bold curve in the ratio profile represents a mean of four chromosomes and is surrounded by thin curves indicating one standard deviation. The central vertical line (broken) indicates a ratio value of 1 (no change), and the vertical lines next to it (dotted) indicate a ratio of 0.5 (left) and 2.0 (right). In chromosomes where the non-invasive tumor 335 used for comparison showed alterations in DNA content, the ratio profile of that chromosome is shown to the right of the invasive tumor profile. The colored bars represent one gene each, identified by the running numbers above the bars (the name of the gene can be seen at [www.MDL.DK/sdata.html](http://www.MDL.DK/sdata.html)). The bars indicate the purported location of the gene, and the colors indicate the expression level of the gene in the invasive tumor compared with the non-invasive counterpart: >2-fold increase (black), >2-fold decrease (blue), no significant change (orange). The bar to the far right, entitled 'Expression' shows the resulting change in expression along the chromosome; the colors indicate that at least half of the genes were up-regulated (black), at least half of the genes down-regulated (blue), or more than half of the genes are unchanged (orange). If a gene was absent in one of the samples and present in another, it was regarded as more than a 2-fold change. A 2-fold level was chosen as this corresponded to one standard deviation in a double determination of ~1800 genes. Centromeres and heterochromatic regions were excluded from data analysis.

grade I and II, respectively, tumors 733 and 827 were staged as pT1 (invasive into submucosa), 733 was staged as solid, and 827 was staged as papillary, both grade III.

**mRNA Preparation**—Tissue biopsies, obtained fresh from surgery, were embedded immediately in a sodium-guanidinium thiocyanate solution and stored at  $-80^{\circ}\text{C}$ . Total RNA was isolated using the RNeasy B RNA isolation method (WAK-Chemie Medical GmbH). poly(A)<sup>+</sup> RNA was isolated by an oligo(dT) selection step (Oligotex mRNA kit; Qiagen).

**cRNA Preparation**—1  $\mu\text{g}$  of mRNA was used as starting material. The first and second strand cDNA synthesis was performed using the SuperScript<sup>®</sup> choice system (Invitrogen) according to the manufacturer's instructions but using an oligo(dT) primer containing a T7 RNA polymerase binding site. Labeled cRNA was prepared using the ME-GAscript<sup>®</sup> *in vitro* transcription kit (Ambion). Biotin-labeled CTP and

UTP (Enzo) was used, together with unlabeled NTPs in the reaction. Following the *in vitro* transcription reaction, the unincorporated nucleotides were removed using RNeasy-columns (Qiagen).

**Array Hybridization and Scanning**—Array hybridization and scanning was modified from a previous method (13). 10  $\mu\text{g}$  of cRNA was fragmented at  $94^{\circ}\text{C}$  for 35 min in buffer containing 40 mM Tris acetate, pH 8.1, 100 mM KOAc, 30 mM MgOAc. Prior to hybridization, the fragmented cRNA in a 6 $\times$  SSPE-T hybridization buffer (1 M NaCl, 10 mM Tris, pH 7.6, 0.005% Triton), was heated to  $95^{\circ}\text{C}$  for 5 min, subsequently cooled to  $40^{\circ}\text{C}$ , and loaded onto the Affymetrix probe array cartridge. The probe array was then incubated for 16 h at  $40^{\circ}\text{C}$  at constant rotation (60 rpm). The probe array was exposed to 10 washes in 6 $\times$  SSPE-T at  $25^{\circ}\text{C}$  followed by 4 washes in 0.5 $\times$  SSPE-T at  $50^{\circ}\text{C}$ . The biotinylated cRNA was stained with a streptavidin-phycoerythrin conjugate, 10  $\mu\text{g}/\text{ml}$  (Molecular Probes) in 6 $\times$  SSPE-T

# Gene Copy Numbers, Transcripts, and Protein Levels

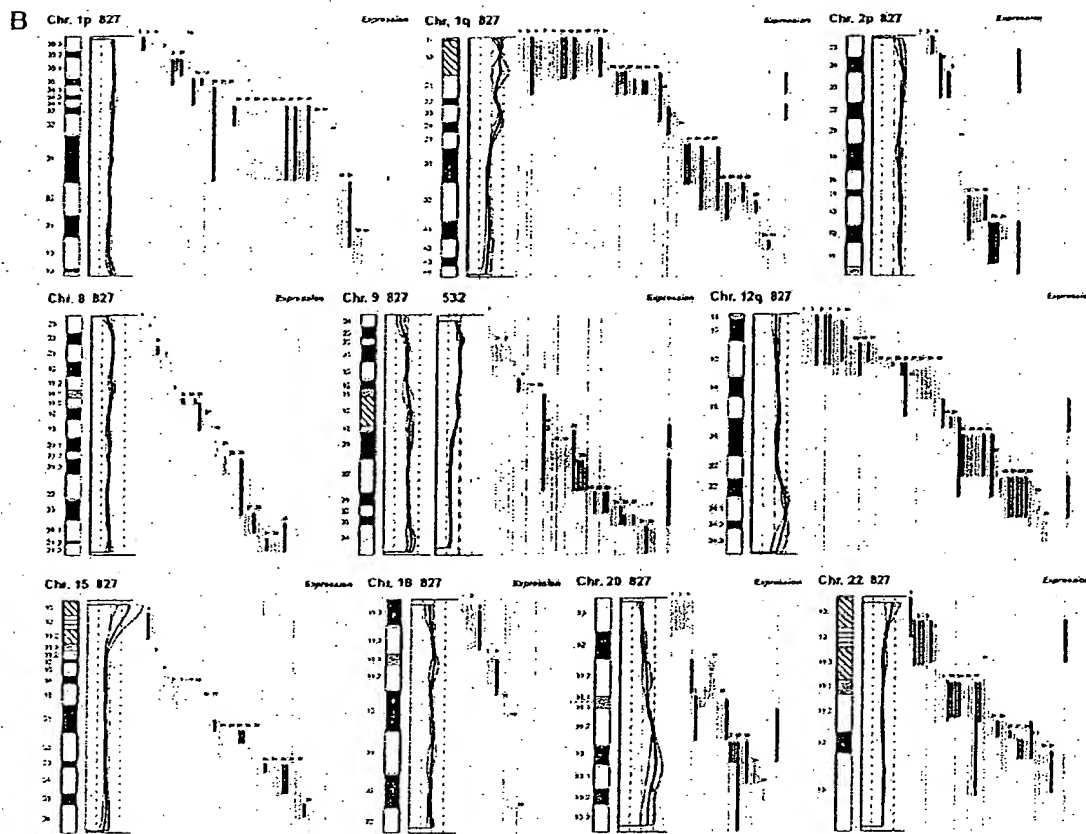


FIG. 1—continued

for 30 min at 25 °C followed by 10 washes in 6× SSPE-T at 25 °C. The probe arrays were scanned at 560 nm using a confocal laser scanning microscope (made for Affymetrix by Hewlett-Packard). The readings from the quantitative scanning were analyzed by Affymetrix gene expression analysis software.

**Microsatellite Analysis.**—Microsatellite Analysis was performed as described previously (14). Microsatellites were selected by use of [www.ncbi.nlm.nih.gov/genemap98](http://www.ncbi.nlm.nih.gov/genemap98), and primer sequences were obtained from the genome data base at [www.gdb.org](http://www.gdb.org). DNA was extracted from tumor and blood and amplified by PCR in a volume of 20  $\mu$ l for 35 cycles. The amplicons were denatured and electrophoresed for 3 h in an ABI Prism 377. Data were collected in the Gene Scan program for fragment analysis. Loss of heterozygosity was defined as less than 33% of one allele detected in tumor amplicons compared with blood.

**Proteomic Analysis.**—TCCs were minced into small pieces and homogenized in a small glass homogenizer in 0.5 ml of lysis solution. Samples were stored at -20 °C until use. The procedure for 2D gel electrophoresis has been described in detail elsewhere (15, 16). Gels were stained with silver nitrate and/or Coomassie Brilliant Blue. Proteins were identified by a combination of procedures that included microsequencing, mass spectrometry, two-dimensional gel Western immunoblotting, and comparison with the master two-dimensional gel image of human keratinocyte proteins; see [biobase.dk/cgi-bin/celfs](http://biobase.dk/cgi-bin/celfs).

**CGH.**—Hybridization of differentially labeled tumor and normal DNA to normal metaphase chromosomes was performed as described previously (10). Fluorescein-labeled tumor DNA (200 ng), Texas Red-

labeled reference DNA (200 ng), and human Cot-1 DNA (20  $\mu$ g) were denatured at 37 °C for 5 min and applied to denatured normal metaphase slides. Hybridization was at 37 °C for 2 days. After washing, the slides were counterstained with 0.15  $\mu$ g/ml 4,6-diamidino-2-phenylindole in an anti-fade solution. A second hybridization was performed for all tumor samples using fluorescein-labeled reference DNA and Texas Red-labeled tumor DNA (inverse labeling) to confirm the aberrations detected during the initial hybridization. Each CGH experiment also included a normal control hybridization using fluorescein- and Texas Red-labeled normal DNA. Digital image analysis was used to identify chromosomal regions with abnormal fluorescence ratios, indicating regions of DNA gains and losses. The average green:red fluorescence intensity ratio profiles were calculated using four images of each chromosome (eight chromosomes total) with normalization of the green:red fluorescence intensity ratio for the entire metaphase and background correction. Chromosome identification was performed based on 4,6-diamidino-2-phenylindole banding patterns. Only images showing uniform high intensity fluorescence with minimal background staining were analyzed. All centromeres, p arms of acrocentric chromosomes, and heterochromatic regions were excluded from the analysis.

## RESULTS

**Comparative Genomic Hybridization.**—The CGH analysis identified a number of chromosomal gains and losses in the

## Gene Copy Numbers, Transcripts, and Protein Levels

TABLE I  
Correlation between alterations detected by CGH and by expression monitoring

Top, CGH used as independent variable (if CGH alteration – what expression ratio was found); bottom, altered expression used as independent variable (if expression alteration – what CGH deviation was found).

CGH alterations	Tumor 733 vs. 335		Concordance	CGH alterations	Tumor 827 vs. 532		Concordance
	Expression change clusters				Expression change clusters		
13 Gain	10 Up-regulation		77%	10 Gain	8 Up-regulation		80%
	0 Down-regulation				0 Down-regulation		
	3 No change				2 No change		
10 Loss	1 Up-regulation		50%	12 Loss	3 Up-regulation		17%
	5 Down-regulation				2 Down regulation		
	4 No change				7 No change		
Expression change clusters	Tumor 733 vs. 335		Concordance	Expression change clusters	Tumor 827 vs. 532		Concordance
	CGH alterations				CGH alterations		
16 Up-regulation	11 Gain		69%	17 Up-regulation	10 Gain		59%
	2 Loss				5 Loss		
	3 No change				2 No change		
21 Down-regulation	1 Gain		38%	9 Down-regulation	0 Gain		33%
	8 Loss				3 Loss		
	12 No change				6 No change		
15.No change	3 Gain		60%	21.No change	1 Gain		81%
	3 Loss				3 Loss		
	9 No change				17 No change		

two invasive tumors (stage pT1, TCCs 733 and 827), whereas the two non-invasive papillomas (stage pTa, TCCs 335 and 532) showed only 9p–, 9q22–q33–, and X–, and 7+, 9q–, and Y–, respectively. Both invasive tumors showed changes (1q22–24+, 2q14.1–qter–, 3q12–q13.3–, 6q12–q22–, 9q34+, 11q12–q13+, 17+, and 20q11.2–q12+) that are typical for their disease stage, as well as additional alterations, some of which are shown in Fig. 1. Areas with gains and losses deviated from the normal copy number to some extent, and the average numerical deviation from normal was 0.4-fold in the case of TCC 733 and 0.3-fold for TCC 827. The largest changes, amounting to at least a doubling of chromosomal content, were observed at 1q23 in TCC 733 (Fig. 1A) and 20q12 in TCC 827 (Fig. 1B).

**mRNA Expression in Relation to DNA Copy Number**—The mRNA levels from the two invasive tumors (TCCs 827 and 733) were compared with the two non-invasive counterparts (TCCs 532 and 335). This was done in two separate experiments in which we compared TCCs 733 to 335 and 827 to 532, respectively, using two different scaling settings for the arrays to rule out scaling as a confounding parameter. Approximately 1,800 genes that yielded a signal on the arrays were searched in the Unigene and Genemap data bases for chromosomal location, and those with a known location (1096) were plotted as bars covering their purported locus. In that way it was possible to construct a graphic presentation of DNA copy number and relative mRNA levels along the individual chromosomes (Fig. 1).

For each mRNA a ratio was calculated between the level in the invasive versus the non-invasive counterpart. Bars, which represent chromosomal location of a gene, were color-coded according to the expression ratio, and only differences larger

than 2-fold were regarded as informative (Fig. 1). The density of genes along the chromosomes varied, and areas containing only one gene were excluded from the calculations. The resolution of the CGH method is very low, and some of the outlier data may be because of the fact that the boundaries of the chromosomal aberrations are not known at high resolution.

Two sets of calculations were made from the data. For the first set we used CGH alterations as the independent variable and estimated the frequency of expression alterations in these chromosomal areas. In general, areas with a strong gain of chromosomal material contained a cluster of genes having increased mRNA expression. For example, both chromosomes 1q21–q25, 2p and 9q, showed a relative gain of more than 100% in DNA copy number that was accompanied by increased mRNA expression levels in the two tumor pairs (Fig. 1). In most cases, chromosomal gains detected by CGH were accompanied by an increased level of transcripts in both TCCs 733 (77%) and 827 (80%) (Table I, top). Chromosomal losses, on the other hand, were not accompanied by decreased expression in several cases, and were often registered as having unaltered RNA levels (Table I, top). The inability to detect RNA expression changes in these cases was not because of fewer genes mapping to the lost regions (data not shown).

In the second set of calculations we selected expression alterations above 2-fold as the independent variable and estimated the frequency of CGH alterations in these areas. As above, we found that increased transcript expression correlated with gain of chromosomal material (TCC 733, 69% and TCC 827, 59%), whereas reduced expression was often detected in areas with unaltered CGH ratios (Table I, bottom). Furthermore, as a control we looked at areas with no alter-

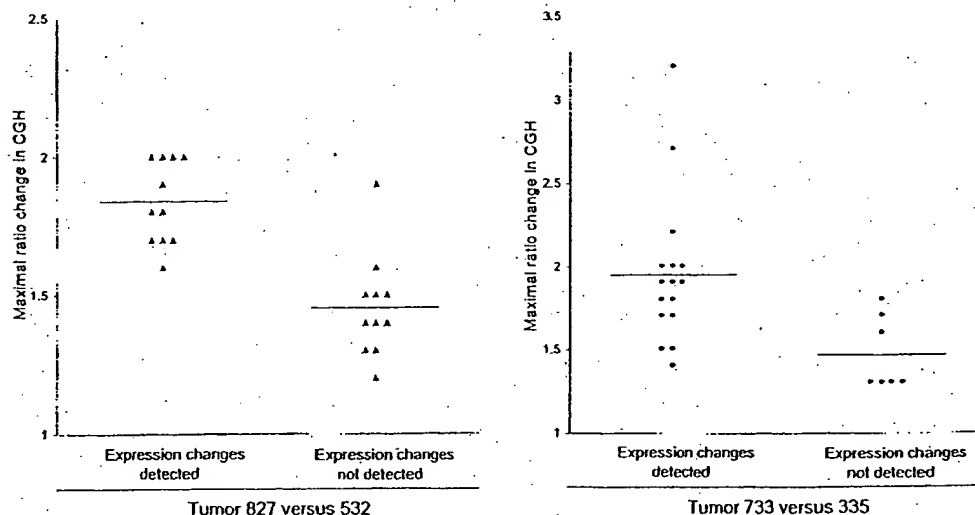


FIG. 2. Correlation between maximum CGH aberration and the ability to detect expression change by oligonucleotide array monitoring. The aberration is shown as a numerical -fold change in ratio between invasive tumors 827 (▲) and 733 (◆) and their non-invasive counterparts 532 and 335. The expression change was taken from the Expression line to the right in Fig. 1, which depicts the resulting expression change for a given chromosomal region. At least half of the mRNAs from a given region have to be either up- or down-regulated to be scored as an expression change. All chromosomal arms in which the CGH ratio plus or minus one standard deviation was outside the ratio value of one were included.

ation in expression. No alteration was detected by CGH in most of these areas (TCC 733, 60% and TCC 827, 81%; see Table I, bottom). Because the ability to observe reduced or increased mRNA expression clustering to a certain chromosomal area clearly reflected the extent of copy number changes, we plotted the maximum CGH aberrations in the regions showing CGH changes against the ability to detect a change in mRNA expression as monitored by the oligonucleotide arrays (Fig. 2). For both tumors TCC 733 ( $p < 0.015$ ) and TCC 827 ( $p < 0.00003$ ) a highly significant correlation was observed between the level of CGH ratio change (reflecting the DNA copy number) and alterations detected by the array based technology (Fig. 2). Similar data were obtained when areas with altered expression were used as independent variables. These areas correlated best with CGH when the CGH ratio deviated 1.6- to 2.0-fold (Table I, bottom) but mostly did not at lower CGH deviations. These data probably reflect that loss of an allele may only lead to a 50% reduction in expression level, which is at the cut-off point for detection of expression alterations. Gain of chromosomal material can occur to a much larger extent.

**Microsatellite-based Detection of Minor Areas of Losses**—In TCC 733, several chromosomal areas exhibiting DNA amplification were preceded or followed by areas with a normal CGH but reduced mRNA expression (see Fig. 1, TCC 733 chromosome 1q32, 2p21, and 7q21 and q32, 9q34, and 10q22). To determine whether these results were because of undetected loss of chromosomal material in these regions or

because of other non-structural mechanisms regulating transcription, we examined two microsatellites positioned at chromosome 1q25-32 and two at chromosome 2p22. Loss of heterozygosity (LOH) was found at both 1q25 and at 2p22 indicating that minor deleted areas were not detected with the resolution of CGH (Fig. 3). Additionally, chromosome 2p in TCC 733 showed a CGH pattern of gain/no change/gain of DNA that correlated with transcript increase/decrease/increase. Thus, for the areas showing increased expression there was a correlation with the DNA copy number alterations (Fig. 1A). As indicated above, the mRNA decrease observed in the middle of the chromosomal gain was because of LOH, implying that one of the mechanisms for mRNA down-regulation may be regions that have undergone smaller losses of chromosomal material. However, this cannot be detected with the resolution of the CGH method.

In both TCC 733 and TCC 827, the telomeric end of chromosome 11p showed a normal ratio in the CGH analysis; however, clusters of five and three genes, respectively, lost their expression. Two microsatellites (D11S1760, D11S922) positioned close to MUC2, IGF2, and cathepsin D indicated LOH as the most likely mechanism behind the loss of expression (data not shown).

A reduced expression of mRNA observed in TCC 733 at chromosomes 3q24, 11p11, 12p12.2, 12q21.1, and 16q24 and in TCC 827 at chromosome 11p15.5, 12p11, 15q11.2, and 18q12 was also examined for chromosomal losses using microsatellites positioned as close as possible to the gene loci

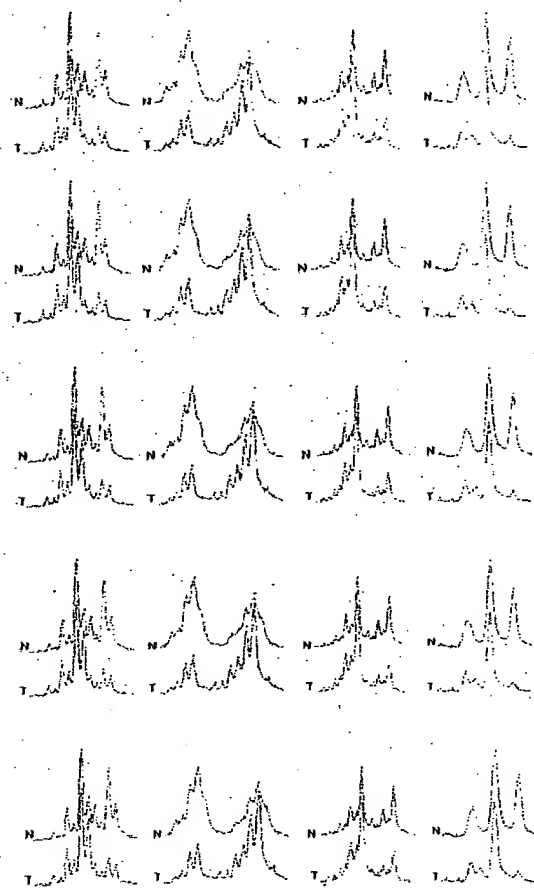


FIG. 3. Microsatellite analysis of loss of heterozygosity. Tumor 733 showing loss of heterozygosity at chromosome 1q25, detected (a) by D1S215 close to HLA class I histocompatibility antigen (gene number 38 in Fig. 1), (b) by D1S2735 close to cathepsin E (gene number 41 in Fig. 1), and (c) at chromosome 2p23 by D2S2251 close to general  $\beta$ -spectrin (gene number 11 on Fig. 1) and of (d) tumor 827 showing loss of heterozygosity at chromosome 18q12 by S18S1118 close to mitochondrial 3-oxoacyl-coenzyme A thiolase (gene number 12 in Fig. 1). The upper curves show the electropherogram obtained from normal DNA from leukocytes (N), and the lower curves show the electropherogram from tumor DNA (T). In all cases one allele is partially lost in the tumor amplicon.

showing reduced mRNA transcripts. Only the microsatellite positioned at 18q12 showed LOH (Fig. 3), suggesting that transcriptional down-regulation of genes in the other regions may be controlled by other mechanisms.

**Relation between Changes in mRNA and Protein Levels—**2D-PAGE analysis, in combination with Coomassie Brilliant Blue and/or silver staining, was carried out on all four tumors using fresh biopsy material. 40 well resolved abundant known proteins migrating in areas away from the edges of the pH

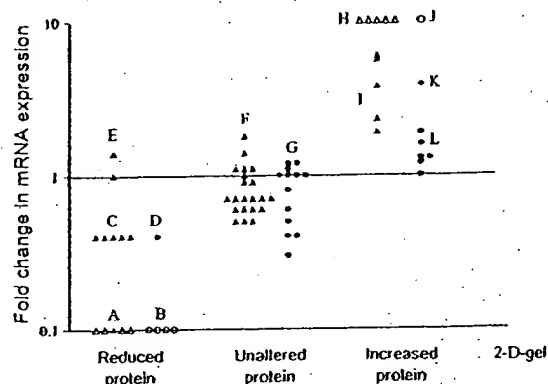


FIG. 4. Correlation between protein levels as judged by 2D-PAGE and transcript ratio. For comparison proteins were divided in three groups, unaltered in level or up- or down-regulated (horizontal axis). The mRNA ratio as determined by oligonucleotide arrays was plotted for each gene (vertical axis).  $\Delta$ , mRNAs that were scored as present in both tumors used for the ratio calculation;  $\Delta$ , mRNAs that were scored as absent in the invasive tumors (along horizontal axis) or as absent in non-invasive reference (top of figure). Two different scalings were used to exclude scaling as a confounder, TCCs 827 and 532 ( $\Delta\Delta$ ) were scaled with background suppression, and TCCs 733 and 335 ( $\bullet\bullet$ ) were scaled without suppression. Both comparisons showed highly significant ( $p < 0.005$ ) differences in mRNA ratios between the groups. Proteins shown were as follows: Group A (from left), phosphoglucomutase 1, glutathione transferase class  $\mu$  number 4, fatty acid-binding protein homologue, cytochrome 15, and cytochrome 13; B (from left), fatty acid-binding protein homologue, 28-kDa heat shock protein, cytochrome 13, and calyculin; C (from left),  $\alpha$ -enolase, hnRNP B1, 28-kDa heat shock protein, 14-3-3- $\epsilon$ , and pre-mRNA splicing factor; D, mesothelial keratin K7 (type II); E (from top), glutathione S-transferase- $\pi$  and mesothelial keratin K7 (type II); F (from top and left), adenyl cyclase-associated protein, E-cadherin, keratin 19, calgizzarin, phosphoglycerate mutase, annexin IV, cytoskeletal  $\gamma$ -actin, hnRNP A1, integral membrane protein calnexin (IP90), hnRNP H, brain-type clathrin light chain- $\alpha$ , hnRNP F, 70-kDa heat shock protein, heterogeneous nuclear ribonucleoprotein A/B, translationally controlled tumor protein, liver glyceraldehyde-3-phosphate dehydrogenase, keratin 8, aldehyde reductase, and Na,K-ATPase  $\beta$ -1 subunit; G, (from top and left), TCP20, calgizzarin, 70-kDa heat shock protein, calnexin, hnRNP H, cytochrome 15, ATP synthase, keratin 19, triosephosphate isomerase, hnRNP F, liver glyceraldehyde-3-phosphate dehydrogenase, glutathione S-transferase- $\pi$ , and keratin 8; H (from left), plasma gelsolin, autoantigen calreticulin, thioredoxin, and NAD $^{+}$ -dependent 15-hydroxyprostaglandin dehydrogenase; I (from top), prollyl 4-hydroxylase  $\beta$ -subunit, cytochrome 20, cytochrome 17, prothionin, and fructose 1,6-bisphosphatase; J annexin II; K, annexin IV; L (from top and left), 90-kDa heat shock protein, prollyl 4-hydroxylase  $\beta$ -subunit,  $\alpha$ -enolase, GRP 78, cyclophilin, and cofillin.

gradient, and having a known chromosomal location, were selected for analysis in the TCC pair 827/532. Proteins were identified by a combination of methods (see "Experimental Procedures"). In general there was a highly significant correlation ( $p < 0.005$ ) between mRNA and protein alterations (Fig. 4). Only one gene showed disagreement between transcript alteration and protein alteration. Except for a group of cyto-

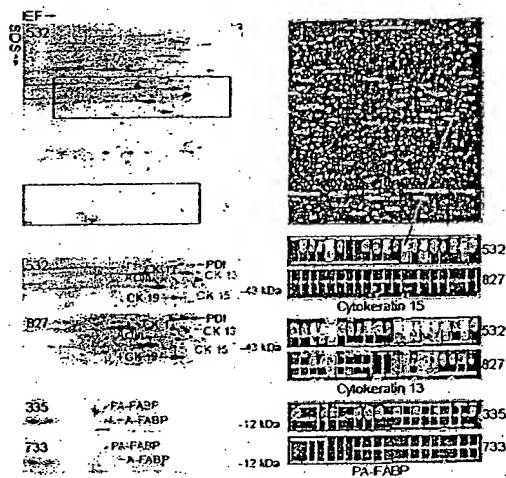


Fig. 5. Comparison of protein and transcript levels in invasive and non-invasive TCCs. The upper part of the figure shows a 2D gel (left) and the oligonucleotide array (right) of TCC 532. The red rectangles on the upper gel highlight the areas that are compared below. Identical areas of 2D gels of TCCs 532 and 827 are shown below. Clearly, cytokeratins 13 and 15 are strongly down-regulated in TCC 827 (red annotation). The tile on the array containing probes for cytokeratin 15 is enlarged below the array (red arrow) from TCC 532 and is compared with TCC 827. The upper row of squares in each tile corresponds to perfect match probes; the lower row corresponds to mismatch probes containing a mutation (used for correction for unspecific binding). Absence of signal is depicted as black, and the higher the signal the lighter the color. A high transcript level was detected in TCC 532 (6151 units) whereas a much lower level was detected in TCC 827 (absence of signals). For cytokeratin 13, a high transcript level was also present in TCC 532 (15659 units), and a much lower level was present in TCC 827 (623 units). The 2D gels at the bottom of the figure (left) show levels of PA-FABP and adipocyte-FABP in TCCs 335 and 733 (invasive), respectively. Both proteins are down-regulated in the invasive tumor. To the right we show the array tiles for the PA-FABP transcript. A medium transcript level was detected in the case of TCC 335 (1277 units) whereas very low levels were detected in TCC 733 (166 units). IEF, isoelectric focusing.

keratins encoded by genes on chromosome 17 (Fig. 5) the analyzed proteins did not belong to a particular family. 26 well focused proteins whose genes had a known chromosomal location were detected in TCCs 733 and 335, and of these 19 correlated ( $p < 0.005$ ) with the mRNA changes detected using the arrays (Fig. 4). For example, PA-FABP was highly expressed in the non-invasive TCC 335 but lost in the invasive counterpart (TCC 733; see Fig. 5). The smaller number of proteins detected in both 733 and 335 was because of the smaller size of the biopsies that were available.

11 chromosomal regions where CGH showed aberrations that corresponded to the changes in transcript levels also showed corresponding changes in the protein level (Table II). These regions included genes that encode proteins that are found to be frequently altered in bladder cancer, namely cytokeratins 17 and 20, annexins II and IV, and the fatty acid-binding proteins PA-FABP and FABP1. Four of these proteins were encoded by genes in chromosome 17q, a frequently amplified chromosomal area in invasive bladder cancers.

#### DISCUSSION

Most human cancers have abnormal DNA content, having lost some chromosomal parts and gained others. The present study provides some evidence as to the effect of these gains and losses on gene expression in two pairs of non-invasive and invasive TCCs using high throughput expression arrays and proteomics, in combination with CGH. In general, the results showed that there is a clear individual regulation of the mRNA expression of single genes, which in some cases was superimposed by a DNA copy number effect. In most cases, genes located in chromosomal areas with gains often exhibited increased mRNA expression, whereas areas showing losses showed either no change or a reduced mRNA expression. The latter might be because of the fact that losses most often are restricted to loss of one allele, and the cut-off point for detection of expression alterations was a 2-fold change, thus being at the border of detection. In several cases, how-

TABLE II  
Proteins whose expression level correlates with both mRNA and gene dose changes

Protein	Chromosomal location	Tumor TCC	CGH alteration	Transcript alteration <sup>a</sup>	Protein alteration
Annexin II	1q21	733	Gain	Abs to Pres <sup>a</sup>	Increase
Annexin IV	2p13	733	Gain	3.9-Fold up	Increase
Cytokeratin 17	17q12-q21	827	Gain	3.8-Fold up	Increase
Cytokeratin 20	17q21.1	827	Gain	5.6-Fold up	Increase
(PA-)FABP	8q21.2	827	Loss	10-Fold down	Decrease
FABP1	9q22	827	Gain	2.3-Fold up	Increase
Plasma gelsolin	9q31	827	Gain	Abs to Pres	Increase
Heat shock protein 28	15q12-q13	827	Loss	2.5-Fold up	Decrease
Prohibitin	17q21	827/733	Gain	3.7-/2.5-Fold up <sup>b</sup>	Increase
Prolyl-4-hydroxyl	17q25	827/733	Gain	5.7-/1.6-Fold up	Increase
hnRNPB1	7p15	827	Loss	2.5-Fold down	Decrease

<sup>a</sup> Abs, absent; Pres, present.

<sup>b</sup> In cases where the corresponding alterations were found in both TCCs 827 and 733 these are shown as 827/733.

## Gene Copy Numbers, Transcripts, and Protein Levels

ever, an increase or decrease in DNA copy number was associated with *de novo* occurrence or complete loss of transcript, respectively. Some of these transcripts could not be detected in the non-invasive tumor but were present at relatively high levels in areas with DNA amplifications in the invasive tumors (e.g. in TCC 733 transcript from cellular ligand of annexin II gene (chromosome 1q21) from absent to 2670 arbitrary units; in TCC 827 transcript from small proline-rich protein 1 gene (chromosome 1q12-q21.1) from absent to 1326 arbitrary units). It may be anticipated from these data that significant clustering of genes with an increased expression to a certain chromosomal area indicates an increased likelihood of gain of chromosomal material in this area.

Considering the many possible regulatory mechanisms acting at the level of transcription, it seems striking that the gene dose effects were so clearly detectable in gained areas. One hypothetical explanation may lie in the loss of controlled methylation in tumor cells (17–19). Thus, it may be possible that in chromosomes with increased DNA copy numbers two or more alleles could be demethylated simultaneously leading to a higher transcription level, whereas in chromosomes with losses the remaining allele could be partly methylated, turning off the process (20, 21). A recent report has documented a ploidy regulation of gene expression in yeast, but in this case all the genes were present in the same ratio (22), a situation that is not analogous to that of cancer cells, which show marked chromosomal aberrations, as well as gene dosage effects.

Several CGH studies of bladder cancer have shown that some chromosomal aberrations are common at certain stages of disease progression, often occurring in more than 1 of 3 tumors. In pTa tumors, these include 9p–, 9q–, 1q+, Y– (2, 6), and in pT1 tumors, 2q–, 11p–, 11q–, 1q+, 5p+, 8q+, 17q+, and 20q+ (2–4, 6, 7). The pTa tumors studied here showed similar aberrations such as 9p– and 9q22–q33– and 9q– and Y–, respectively. Likewise, the two minimal invasive pT1 tumors showed aberrations that are commonly seen at that stage, and TCC 827 had a remarkable resemblance to the commonly seen pattern of losses and gains, such as 1q22–24 amplification (seen in both tumors), 11q14–q22 loss, the latter often linked to 17q+ (both tumors), and 1q+ and 9p–, often linked to 20q+ and 11q13+ (both tumors) (7–9). These observations indicate that the pairs of tumors used in this study exhibit chromosomal changes observed in many tumors, and therefore the findings could be of general importance for bladder cancer.

Considering that the mapping resolution of CGH is of about 20 megabases it is only possible to get a crude picture of chromosomal instability using this technique. Occasionally, we observed reduced transcript levels close to or inside regions with increased copy numbers. Analysis of these regions by positioning heterozygous microsatellites as close as possible to the locus showing reduced gene expression revealed loss of heterozygosity in several cases. It seems likely that multiple and different events occur along each chromosomal

arm and that the use of cDNA microarrays for analysis of DNA copy number changes will reach a resolution that can resolve these changes, as has recently been proposed (2). The outlier data were not more frequent at the boundaries of the CGH aberrations. At present we do not know the mechanism behind chromosomal aneuploidy and cannot predict whether chromosomal gains will be transcribed to a larger extent than the two native alleles. A mechanism as genetic imprinting has an impact on the expression level in normal cells and is often reduced in tumors. However, the relation between imprinting and gain of chromosomal material is not known.

We regard it as a strength of this investigation that we were able to compare invasive tumors to benign tumors rather than to normal urothelium, as the tumors studied were biologically very close and probably may represent successive steps in the progression of bladder cancer. Despite the limited amount of fresh tissue available it was possible to apply three different state of the art methods. The observed correlation between DNA copy number and mRNA expression is remarkable when one considers that different pieces of the tumor biopsies were used for the different sets of experiments. This indicates that bladder tumors are relatively homogenous, a notion recently supported by CGH and LOH data that showed a remarkable similarity even between tumors and distant metastasis (10, 23).

In the few cases analyzed, mRNA and protein levels showed a striking correspondence although in some cases we found discrepancies that may be attributed to translational regulation, post-translational processing, protein degradation, or a combination of these. Some transcripts belong to undertranslated mRNA pools, which are associated with few translationally inactive ribosomes; these pools, however, seem to be rare (24). Protein degradation, for example, may be very important in the case of polypeptides with a short half-life (e.g. signaling proteins). A poor correlation between mRNA and protein levels was found in liver cells as determined by arrays and 2D-PAGE (25), and a moderate correlation was recently reported by Ideker *et al.* (26) in yeast.

Interestingly, our study revealed a much better correlation between gained chromosomal areas and increased mRNA levels than between loss of chromosomal areas and reduced mRNA levels. In general, the level of CGH change determined the ability to detect a change in transcript. One possible explanation could be that by losing one allele the change in mRNA level is not so dramatic as compared with gain of material, which can be rather unlimited and may lead to a severalfold increase in gene copy number resulting in a much higher impact on transcript level. The latter would be much easier to detect on the expression arrays as the cut-off point was placed at a 2-fold level so as not to be biased by noise on the array. Construction of arrays with a better signal to noise ratio may in the future allow detection of lesser than 2-fold alterations in transcript levels, a feature that may facilitate the analysis of the effect of loss of chromosomal areas on transcript levels.

In eleven cases we found a significant correlation between DNA copy number, mRNA expression, and protein level. Four of these proteins were encoded by genes located at a frequently amplified area in chromosome 17q. Whether DNA copy number is one of the mechanisms behind alteration of these eleven proteins is at present unknown and will have to be proved by other methods using a larger number of samples. One factor making such studies complicated is the large extent of protein modification that occurs after translation, requiring immunoidentification and/or mass spectrometry to correctly identify the proteins in the gels.

In conclusion, the results presented in this study exemplify the large body of knowledge that may be possible to gather in the future by combining state of the art techniques that follow the pathway from DNA to protein (26). Here, we used a traditional chromosomal CGH method, but in the future high resolution CGH based on microarrays with many thousand radiation hybrid-mapped genes will increase the resolution and information derived from these types of experiments (2). Combined with expression arrays analyzing transcripts derived from genes with known locations, and 2D gel analysis to obtain information at the post-translational level, a clearer and more developed understanding of the tumor genome will be forthcoming.

**Acknowledgments**—We thank Mie Madsen, Hanne Steen, Inge Lis Thorsen, Hans Lund, Viktor Ørntoft, and Lynn Bjerke for technical help and Thomas Gingeras, Christine Harrington, and Morten Østergaard for valuable discussions.

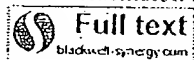
This work was supported by grants from The Danish Cancer Society, the University of Aarhus, Aarhus County, Novo Nordic, the Danish Biotechnology Program, the Frenkels Foundation, the John and Birthe Meyer Foundation, and NCI, National Institutes of Health Grant CA47537. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ To whom correspondence should be addressed: Dept. of Clinical Biochemistry, Molecular Diagnostic Laboratory, Aarhus University Hospital, Skejby, DK-8200 Aarhus N, Denmark. Tel.: 45-89495100/45-86156201 (private); Fax: 45-89496018; E-mail: ornott@kba.sks.au.dk.

## REFERENCES

- Lengauer, C., Kinzler, K. W., and Vogelstein, B. (1998) Genetic instabilities in human cancers. *Nature* 396, 643-649.
- Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D., and Brown, P. O. (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.* 23, 41-46.
- de Cremoux, P., Martin, E. C., Vincent-Salomon, A., Dieras, V., Barbaroux, C., Liva, S., Pouillart, P., Sastre-Garau, X., and Magdelenat, H. (1999) Quantitative PCR analysis of c-erb B-2 (HER2/neu) gene amplification and comparison with p185(HER2/neu) protein expression in breast cancer drill biopsies. *Int. J. Cancer* 83, 157-161.
- Brugier, P. P., Tamimi, Y., Shuring, E., and Schalken, J. (1996) Expression of cyclin D1 and EMS1 in bladder tumors; relationship with chromosome 11q13 amplifications. *Oncogene* 12, 1747-1753.
- Slavc, I., Ellenbogen, R., Jung, W. H., Vawter, G. F., Kretschmar, C., Grier, H., and Korf, B. R. (1990) myc gene amplification and expression in primary human neuroblastoma. *Cancer Res.* 50, 1459-1463.
- Sauter, G., Carroll, P., Moch, H., Kallioniemi, A., Kerschmann, R., Narayan, P., Mihatsch, M. J., and Waldman, F. M. (1995) c-myc copy number gains in bladder cancer detected by fluorescence *in situ* hybridization. *Am. J. Pathol.* 146, 1131-1139.
- Richter, J., Jiang, F., Gorog, J. P., Sartorius, G., Egenter, C., Gasser, T. C., Moch, H., Mihatsch, M. J., and Sauter, G. (1997) Marked genetic differences between stage pTa and stage pT1 papillary bladder cancer detected by comparative genomic hybridization. *Cancer Res.* 57, 2860-2864.
- Richter, J., Boffa, L., Wagner, U., Schraml, P., Gasser, T. C., Moch, H., Mihatsch, M. J., and Sauter, G. (1998) Patterns of chromosomal imbalances in advanced urinary bladder cancer detected by comparative genomic hybridization. *Am. J. Pathol.* 153, 1615-1621.
- Bruch, J., Wöhr, G., Hautmann, R., Mattfeldt, T., Bruderlein, S., Möller, P., Sauter, S., Hameister, H., Vogel, W., and Paiss, T. (1998) Chromosomal changes during progression of transitional cell carcinoma of the bladder and delineation of the amplified interval on chromosome arm 8q. *Genes Chromosomes Cancer* 23, 167-174.
- Hovey, R. M., Chu, L., Balazs, M., De Vries, S., Moore, D., Sauter, G., Carroll, P. R., and Waldman, F. M. (1998) Genetic alterations in primary bladder cancers and their metastases. *Cancer Res.* 58, 3555-3560.
- Simon, R., Burger, H., Brinkschmidt, C., Bocker, W., Hertle, L., and Terpe, H. J. (1998) Chromosomal aberrations associated with invasion in papillary superficial bladder cancer. *J. Pathol.* 185, 345-351.
- Koo, S. H., Kwon, K. C., Ihm, C. H., Jeon, Y. M., Park, J. W., and Sul, C. K. (1999) Detection of genetic alterations in bladder tumors by comparative genomic hybridization and cytogenetic analysis. *Cancer Genet. Cytogenet.* 110, 87-93.
- Wodicka, L., Dong, H., Mittmann, M., Ho, M. H., and Lockhart, D. J. (1997) Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nat. Biotechnol.* 15, 1359-1367.
- Christensen, M., Sunde, L., Bolund, L., and Ornott, T. F. (1999) Comparison of three methods of microsatellite detection. *Scand. J. Clin. Lab. Invest.* 59, 167-177.
- Celis, J. E., Østergaard, M., Basse, B., Celis, A., Lauridsen, J. B., Ratz, G. P., Andersen, I., Hein, B., Wolf, H., Ornott, T. F., and Rasmussen, H. H. (1996) Loss of adipocyte-type fatty acid binding protein and other protein biomarkers is associated with progression of human bladder transitional cell carcinomas. *Cancer Res.* 56, 4782-4790.
- Celis, J. E., Ratz, G., Basse, B., Lauridsen, J. B., and Celis, A. (1994) in *Cell Biology: A Laboratory Handbook* (Celis, J. E., ed) Vol. 3, pp. 222-230, Academic Press, Orlando, FL.
- Ohlsson, R., Tyckö, B., and Sapienza, C. (1998) Monoallelic expression: 'there can only be one'. *Trends Genet.* 14, 435-438.
- Hollander, G. A., Zuklys, S., Morel, C., Mizoguchi, E., Mobisson, K., Simpson, S., Terhorst, C., Wishart, W., Golan, D. E., Bhan, A. K., and Burakoff, S. J. (1998) Monoallelic expression of the interleukin-2 locus. *Science* 279, 2118-2121.
- Brannan, C. I., and Bartolomei, M. S. (1999) Mechanisms of genomic imprinting. *Curr. Opin. Genet. Dev.* 9, 164-170.
- Ohlsson, R., Cui, H., He, L., Pfeifer, S., Malmkumpu, H., Jiang, S., Feinberg, A. P., and Hedborg, F. (1999) Mosaic allelic insulin-like growth factor 2 expression patterns reveal a link between Wilms' tumorigenesis and epigenetic heterogeneity. *Cancer Res.* 59, 3889-3892.
- Cui, H., Hedborg, F., He, L., Nordenskjöld, A., Sandstedt, B., Pfeifer, Ohlsson, S., and Ohlsson, R. (1997) Inactivation of H19, an imprinted and putative tumor repressor gene, is a preneoplastic event during Wilms' tumorigenesis. *Cancer Res.* 57, 4469-4473.
- Galitski, T., Saldanha, A. J., Styles, C. A., Lander, E. S., and Fink, G. R. (1999) Ploidy regulation of gene expression. *Science* 285, 251-254.
- Tsao, J., Yatabe, Y., Mark, I. D., Haiyan, K., Jones, P. A., and Shibata, D. (2000) Bladder cancer genotype stability during clinical progression. *Genes Chromosomes Cancer* 29, 26-32.
- Zong, O., Schummer, M., Hood, L., and Morris, D. R. (1999) Messenger RNA translation state: the second dimension of high-throughput expression screening. *Proc. Natl. Acad. Sci. U. S. A.* 96, 10632-10636.
- Anderson, L., and Selinger, J. (1997) Comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* 18, 533-537.
- Ideker, T., Thorsson, V., Ransh, J. A., Christmas, R., Buhler, J., Eng, J. K., Bumgarner, R., Goodlett, D. R., Aebersold, R., and Hood, L. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 292, 929-934.





**Expression of bcr-abl mRNA in individual chronic myelogenous leukaemia cells as determined by in situ amplification.**

Pachmann K, Zhao S, Schenk T, Kantarjian H, El-Naggar AK, Siciliano MJ, Guo JQ, Arlinghaus RB, Andreeff M.

The University of Texas M.D. Anderson Cancer Center, Department of Molecular Haematology and Therapy, 1515 Holcombe Boulevard, Houston, TX 77030, USA.

We present the results of a novel method developed for evaluation of in situ amplification, a molecular genetic method at the cellular level. Reverse transcription polymerase chain reaction (RT-PCR) was used to study bcr-abl transcript levels in individual cells from patients with chronic myelogenous leukaemia (CML). After hybridizing a fluorochrome-labelled probe to the cell-bound RT-PCR product, bcr-abl mRNA-positive cells were determined using image analysis. A dilution series of bcr-abl-positive BV173 into normal cells showed a good correlation between expected and actual values. In 25 CML samples, the percentage of in situ PCR-positive cells showed an excellent correlation with cytogenetic results ( $r = 0.94$ ,  $P < 0.0001$ ), interphase fluorescence in situ hybridization (FISH) ( $r = 0.95$ ,  $P = 0.001$ ) and hypermetaphase FISH ( $r = 0.81$ ,  $P < 0.001$ ). The fluorescence intensity was higher in residual CML cells after interferon (IFN) treatment than in newly diagnosed patients ( $P = 0.004$ ), and was highest in late-stage CML resistant to IFN therapy and lowest in CML blast crisis ( $P = 0.001$ ). Mean fluorescence values correlated with bcr-abl protein levels, as determined by Western blot analysis ( $r = 0.62$ ). Laser scanning cytometry allowing automated analysis of large numbers of cells confirmed the results. Thus, fluorescence in situ PCR provides a novel and quantitative approach for monitoring tumour load and bcr-abl transcript levels in CML.

PMID: 11260080 [PubMed - indexed for MEDLINE]

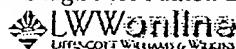
**Cell localization and regulation of expression of cytochrome P450 1A1 and 2B1 in rat lung after induction with 3-methylcholanthrene using mRNA hybridization and immunohistochemistry.**

Pairon JC, Trabelsi N, Buard A, Fleury-Feith J, Bachelet CM, Poron F, Beaune P, Brochard P, Laurent P.

INSERM Unite 139, Hopital Henri Mondor, Creteil, France.

In order to characterize the response of various pulmonary cell types to polycyclic aromatic hydrocarbons, the expression of cytochrome P450 (CYP) 1A1 and 2B1 mRNA in the lung of rats, with or without induction by 3-methylcholanthrene (3MC), was analyzed by in situ hybridization using appropriate 35S-labeled riboprobes. The expression of the corresponding proteins was investigated immunohistochemically. Following induction with 3MC, the kinetics of mRNA expression differed considerably between Clara cells and type II pneumocytes and venous endothelial cells. In Clara cells, mRNA expression was detected as early as 1 h after induction, peaked between 2 and 4 h, and was completely undetectable at 14 h. In contrast, venous endothelial cells and type II pneumocytes exhibited permanent mRNA expression of CYP 1A1 in 3MC-pretreated rats. These kinetic results explain the striking absence of correlation between mRNA and protein expression observed in Clara cells 24 h after the end of the induction protocol, as these cells exhibited intense protein expression with no mRNA. In contrast, a good correlation was observed for mRNA and protein expression of CYP 2B1, with similar expressions for Clara cells and type II pneumocytes, but no expression in endothelial cells. This study clearly distinguished the regulation of CYP 1A1 expression in the rat lung from that described in the liver. The differences observed in the various lung cell types, whatever the post-transcriptional mechanisms involved, emphasize that studies must be performed at the cellular level in order to understand the specific response to xenobiotics, not only of this organ as a whole but also of its various anatomic structures.

PMID: 7917307 [PubMed - indexed for MEDLINE]



# **Correlative immunohistochemical and reverse transcriptase polymerase chain reaction analysis of somatostatin receptor type 2 in neuroendocrine tumors of the lung.**

**Papotti M, Croce S, Macri L, Funaro A, Pecchioni C, Schindler M, Bussolati G.**

Department of Biomedical Sciences and Oncology, University of Turin, Italy.

Somatostatin receptors type 2 (sst2) have been frequently detected in neuroendocrine tumors and bind somatostatin analogues, such as octreotide, with high affinity. Receptor autoradiography, specific mRNA detection and, more recently, antisst2 polyclonal antibodies are currently employed to reveal sst2. The aim of the present study was to investigate by three different techniques the presence of sst2 in a series of 26 neuroendocrine tumors of the lung in which fresh frozen tissue and paraffin sections were available. It was possible, therefore, to compare, in individual cases, RNA analysis studied by reverse transcriptase polymerase chain reaction (RT-PCR), in situ hybridization (ISH), and immunohistochemistry. A series of 20 nonneuroendocrine lung carcinoma samples served as controls. RT-PCR was positive for sst2 in 22 of 26 samples, including 15 of 15 typical carcinoids, 5 of 6 atypical carcinoids, and 2 of 5 small-cell carcinomas. The sst2 mRNA signal obtained by RT-PCR was strong in the majority (87%) of typical carcinoids and of variable intensity in atypical carcinoids and small-cell carcinomas. A weakly positive signal was observed in 5 of 20 control samples. In immunohistochemistry, two different antibodies (anti-sst2) were employed, including a monoclonal antibody, generated in the Department of Pathology, University of Turin. In the majority of samples a good correlation between sst2 mRNA (as detected by RT-PCR) and sst2 protein expression (as detected by immunohistochemistry) was observed. However, one atypical carcinoid and one small-cell carcinoma had focal immunostaining but no RT-PCR signal. ISH performed in selected samples paralleled the results obtained with the other techniques. A low sst2 expression was associated with high grade neuroendocrine tumors and with aggressive behavior. It is concluded that 1) neuroendocrine tumors of the lung express sst2, and there is a correlation between the mRNA amount and the degree of differentiation; 2) immunohistochemistry and ISH are reliable tools to demonstrate sst2 in these tumors; and 3) sst2 identification in tissue sections may provide information on the diagnostic or therapeutic usefulness of somatostatin analogues in individual patients with neuroendocrine tumors.

PMID: 10718213 [PubMed - indexed for MEDLINE]

Volume 9, Number 1, March 2000

# **Diagnostic Molecular Pathology**

WESTON LIBRARY

MAR 07 2000

35126 CLINICAL SCIENCE CENTER  
600 HIGHLAND AVE MADISON WI 53792

**Co-Editors**

**Ronald A. DeLeellis, M.D.**

**Robert J. Moline, M.D.**



LIPPINCOTT WILLIAMS & WILKINS

# Diagnostic Molecular Pathology

Volume 9 □ Number 1 □ March 2000

## ORIGINAL ARTICLES

- 1 Strong Association of SYT-SSX Fusion Type and Morphologic Epithelial Differentiation in Synovial Sarcoma  
Cristina R. Antonescu, Akira Kawai, Denis H. Leung, Fulvio Lonardo, James M. Woodruff, John H. Healey, and Marc Ladanyi
- 9 Clinical Relevance of Molecular Diagnosis in Childhood Rhabdomyosarcoma  
Ana Tobar, Smadar Avigad, Meira Zoldan, Celia Mor, Yakov Goshen, and Rina Zaizov
- 14 Accumulation of Chromosomal Imbalances From Intraductal Proliferative Lesions to Adjacent In Situ and Invasive Ductal Breast Cancer  
Michaela M. Aubele, Margaret C. Cummings, Anita E. Mattis, Horst F. Zitzelsberger, Axel K. Walch, Markus Kremer, Heinz Höfler, and Martin Werner
- 20 Routine Analysis of p53 Mutation in Clinical Breast Tumor Specimens Using Fluorescence-Based Polymerase Chain Reaction and Single Strand Conformation Polymorphism  
Barry Iacopetta, Hany Elsaleh, Fabienne Grieu, David Joseph, Greg Sterrett, and Peter Robbins
- 26 Tumor-Associated Overexpression of the Soluble T1-S Receptor in Lymph Node-Negative Breast Cancer  
Anne Katrin Werenskiold, Dieter Prechtel, Nadia Harbeck, and Heinz Höfler

(continued on next page)

Listed in *Index Medicus*, *Current Awareness in Biological Sciences*, *EMBASE/Excerpta Medica*,  
*Current Contents/Life Sciences*, and *Science Citation Index*

*Diagnostic Molecular Pathology* (ISSN 1052-9551) is published four times per year in March, June, September, and December by Lippincott Williams & Wilkins, Inc., 12107 Insurance Way, Hagerstown, MD 21740. Business offices are located at 530 Walnut Street, Philadelphia, PA 19106-3621. Printed in the U.S.A. Periodicals postage paid at Hagerstown, MD, and at additional mailing offices.

Copyright © 2000 by Lippincott Williams & Wilkins, Inc. All rights reserved.

Address for subscription information, orders, or changes of address: (except Japan) 12107 Insurance Way, Hagerstown, MD 21740, or call 1-800-638-3030; in Maryland, call collect 301-714-2300. In Japan, contact Igaku-Shoin, Ltd., 1-28-36 Hongo, Bunkyo-ku, Tokyo 113, Japan; phone: 81-3-3817-5675; fax: 81-3-3815-6776.

Annual subscription rates: U.S.: \$143.00 individual, \$309.00 institution; Canada and Mexico: \$165.00 individual, \$232.00 institution. (The Canadian GST Tax of 7% will be added to the subscription price of all orders shipped to Canada. The Lippincott Williams & Wilkins, Inc. GST Identification No. is 895524239.) Canada Post International Publications Mail Product Sales Agreement No. 0616168; all other countries (except Japan): \$171.00 individual, \$258.00 institution. (Prices outside North America include \$6.00 for air freight shipping; air freight delivery occurs within 7-21 days worldwide.) International subscriptions must be prepaid. Single copies, when available, may be ordered from the publisher. Single copies \$38.00. Prices are subject to change without notice. Copies will be replaced without charge if the publisher receives a request within 90 days of the mailing date, both in the U.S. and worldwide.

Postmaster: Send changes of address to *Diagnostic Molecular Pathology*, P.O. Box 1550, Hagerstown, MD 21740.

WESTON LIBRARY  
MAR 07 2000  
JBI20 CLINICAL SCIENCE CENTER  
800 HIGHLAND AVE MADISON WI 53792

This material may be protected by Copyright law (Title 17 U.S. Code)

## Correlative Immunohistochemical and Reverse Transcriptase Polymerase Chain Reaction Analysis of Somatostatin Receptor Type 2 in Neuroendocrine Tumors of the Lung

Mauro Papotti, M.D., Sabrina Croce, M.D., Luigia Macri, M.D.,  
Aola Funaro, Ph.D., Carla Pecchioni, Marcus Schindler, M.D., and  
Gianni Bussolati, M.D., F.R.C.Path.

Somatostatin receptors type 2 (sst2) have been frequently detected in neuroendocrine tumors and bind somatostatin analogues, such as octreotide, with high affinity. Receptor autoradiography, specific mRNA detection and, more recently, anti-sst2 polyclonal antibodies are currently employed to reveal sst2. The aim of the present study was to investigate by three different techniques the presence of sst2 in a series of 26 neuroendocrine tumors of the lung in which fresh frozen tissue and paraffin sections were available. It was possible, therefore, to compare, in individual cases, RNA analysis studied by reverse transcriptase polymerase chain reaction (RT-PCR), *in situ* hybridization (ISH), and immunohistochemistry. A series of 20 nonneuroendocrine lung carcinoma samples served as controls. RT-PCR was positive for sst2 in 22 of 26 samples, including 15 of 15 typical carcinoids, 5 of 6 atypical carcinoids, and 2 of 5 small-cell carcinomas. The sst2 mRNA signal obtained by RT-PCR was strong in the majority (87%) of typical carcinoids and of variable intensity in atypical carcinoids and small-cell carcinomas. A weakly positive signal was observed in 5 of 20 control samples. In immunohistochemistry, two different antibodies (anti-sst2) were employed, including a monoclonal antibody, generated in the Department of Pathology, University of Turin. In the majority of samples a good correlation between sst2 mRNA (as detected by RT-PCR) and sst2 protein expression (as detected by immunohistochemistry) was observed. However, one atypical carcinoid and one small-cell carcinoma had focal immunostaining but no RT-PCR signal. ISH performed in selected samples paralleled the results obtained with the other techniques. A low sst2 expression was associated with

high grade neuroendocrine tumors and with aggressive behavior. It is concluded that 1) neuroendocrine tumors of the lung express sst2, and there is a correlation between the mRNA amount and the degree of differentiation; 2) immunohistochemistry and ISH are reliable tools to demonstrate sst2 in these tumors; and 3) sst2 identification in tissue sections may provide information on the diagnostic or therapeutic usefulness of somatostatin analogues in individual patients with neuroendocrine tumors.

**Key Words:** Neuroendocrine—Lung—Tumors—Somatostatin receptors—Immunohistochemistry—Small cell carcinoma—Reverse transcriptase polymerase chain reaction.

*Diagn Mol Pathol* 9(1): 47-57, 2000.

The somatostatin receptor family (sst) includes at least five isoforms that have been recently identified and characterized (18,32,41). The ssts are widely distributed in normal human tissues and in human tumors. Sst type 2 is more commonly detected in neuroendocrine tumors (32,37) and binds the somatostatin analogue octreotide with high affinity.

Sst localization had originally been demonstrated by means of binding assays of radiolabeled somatostatin analogues (20,25,31). Subsequently, specific sst messenger RNA (mRNA) detection was obtained by means of *in situ* hybridization (ISH) and reverse transcriptase polymerase chain reaction (RT-PCR) (14,32,37). Recently, polyclonal antibodies specific for different isoforms of sst were produced and used in immunohistochemistry (10,12,15,18,30,35,36). Given the well-known heterogeneity of neoplastic populations, *in situ* methods (immunohistochemistry and ISH) allow a more definite mapping of the distribution of the receptor in such tissues.

From the Department of Biomedical Sciences and Oncology (M.P., S.C., L.M., A.F., C.P., G.B.), University of Turin, Italy; and Glaxo Institute of Applied Pharmacology (M.S.), University of Cambridge, United Kingdom.

Supported by grants from the Regione Piemonte (grant no. 165 of Research Project DGR 34-23230) and Ministry of University (Rome), Italy.

A.F. is the recipient of a Research Contract by the University of Turin.

Address correspondence and reprint requests to Prof. M. Papotti, Department of Pathology, University of Turin, Via Santena 7, I-10126 Torino.

This is potentially useful for predicting the responsiveness of a given neoplastic cell population to medical treatment with somatostatin analogues, which are used in the clinical setting for both diagnostic and therapeutic purposes with special reference to neuroendocrine tumors.

The spectrum of neuroendocrine tumors of the lung includes well-differentiated neoplasms (so-called typical carcinoids) and poorly differentiated small-cell carcinomas (SCCs). Intermediate forms sharing features of both the aforementioned types also belong to this spectrum (so-called atypical carcinoids or well-differentiated neuroendocrine carcinomas). Finally, large-cell neuroendocrine carcinoma has been identified and included in this tumor group (4,40). The tissue distribution of *sst2* in neuroendocrine tumors of the lung has not been thoroughly characterized, although individual samples of bronchial carcinoids were found to express *sst* (30). SCCs (but not non-small-cell types) were also shown to be *sst2* positive by receptor binding assay (33). Moreover, *sst2* has been detected in *in vitro* cell cultures of human SSC of the lung (39,42). No study on a series of neuroendocrine tumors of the lung including all neuroendocrine lung tumor types has been reported to date.

The aim of this study was therefore to investigate the presence of *sst2* mRNA and protein in a series of 26 neuroendocrine tumors of the lung, employing different technical approaches, such as RT-PCR, ISH, and immunohistochemistry. To this purpose a monoclonal antibody to *sst2* (N-terminal) was generated in the Department of Pathology, University of Turin. The results were then compared and related to the tumor grade and to other clinicopathologic parameters.

## MATERIALS AND METHODS

### Case Series and RNA Extraction

Twenty-six samples of neuroendocrine tumors of the lung, in which fresh frozen tissue was available, were retrieved from the surgical pathology file of the University of Turin, Italy. All samples were reviewed applying currently accepted criteria of classification (4,40), and the neuroendocrine nature was confirmed by positive immunostaining for chromogranin A (CgA) (with or without antigen retrieval) or synaptophysin, and by positive RT-PCR for CgA mRNA. According to the classifications described here, these included 15 well-differentiated neuroendocrine tumors (typical carcinoids), 6 well-differentiated neuroendocrine carcinomas (atypical carcinoids), and 5 SCCs.

A series of 20 non-small-cell lung carcinomas (10 squamous, 9 adenocarcinomas, and 1 large-cell anaplastic) lacking neuroendocrine differentiation, as demonstrated by negative immunohistochemistry and RT-PCR for CgA (1), served as a control group. Clinicopathologic data and follow-up information were obtained for all patients.

For hybridization analysis, total RNA was extracted using the guanidine thiocyanate-cesium chloride method (5). The concentration of RNA was estimated by spectrophotometry, and RNA degradation was assessed by agarose gel electrophoresis, as previously reported (37).

### Reverse Transcriptase Polymerase Chain Reaction for *sst2* and Chromogranin A

Total RNA (2 µg) was first digested, with 10 units of RNase-free DNase (Boehringer, Mannheim, Germany) in a 10-µL solution containing 20 mmol/L MgCl<sub>2</sub>, to avoid DNA contamination. The solution was kept at room temperature for 10 minutes, then heated for 5 minutes at 70°C to inactivate the DNase molecules; 40 pmol/L of oligodeoxythymidine primers (oligo-dT16) were added and the solution was heated again at 70°C for 10 minutes, then chilled on ice to allow the primer hybridization. The resulting solution was reverse transcribed using 100 units of reverse transcriptase (Gibco BRL, Gaithersburg, MD). Complementary DNA (cDNA) was generated in a 50-µL final reaction volume containing 50 mmol/L Tris-HCl pH 8.3, 75 mmol/L KCl, 3 mmol/L MgCl<sub>2</sub>, 10 mmol/L dithiothreitol, 1 mmol/L deoxynucleotide triphosphates (dNTPs), and 20 units of RNasin (Promega, Madison, WI). The solution was heated at 37°C for 90 minutes. Finally, the enzymes were inactivated by heating to 70°C for 10 minutes.

The efficiency of the reverse transcription was determined by performing a PCR reaction having the  $\beta_2$ -microglobulin "housekeeping gene" as a target. PCR was carried out in a 10-µL final reaction volume containing 1 µL of cDNA template, 10 pmol of sense and antisense oligonucleotide primers, 67 mmol/L Tris-HCl pH 8.8, 16 mmol/L (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.01% polysorbate 20, 2 mmol/L dNTPs, 1 mmol/L MgCl<sub>2</sub>, and 0.5 units of Taq polymerase.  $\beta_2$ -Microglobulin, *sst2*, and CgA PCR reactions were performed using the same protocol at the following PCR conditions: 35 cycles, each cycle consisting of denaturation at 94°C for 2 minutes, annealing at 55°C for 1 minute for  $\beta_2$ -microglobulin, at 61°C for *sst2*, and at 68°C for CgA; extension was performed at 72°C for 1 minute. The primers used for RT-PCR (9,11,23,37) are reported in Table 1.

The amplified fragments were run in a 1% agarose gel, containing ethidium bromide. Strict precautions against contamination were undertaken (19) and negative controls (a no-template control and a no-reverse transcriptase control and distilled water to replace the RNA) were included. The RNA extracted from an H716 neuroendocrine colon carcinoma cell line and from a neuroblastoma (37) served as positive controls for CgA and *sst2*, respectively.

### Antibodies

Two different antibodies specific for *sst2* were employed. The first one was a monoclonal antibody raised

TABLE 1. Sequences of primers used for reverse transcriptase polymerase chain reaction

	Size of PCR product (bp)	Position	Study
1) $\beta_2$ -microglobulin sense: 5' ACC CCC ACT GAA AAA GAT GA 3'	120	286-305	Gussow et al. (9)
2) $\beta_2$ -microglobulin antisense: 5' ATG TTC AAA CCT CCA TGA TG 3'		389-408	
3) SSTR2 sense: 5' CAG TCA TGA GCA TCG ACC GA 3'	284	402-421	Sestini et al. (37)
4) SSTR2 antisense: 5' GCA AAG ACA GAT GAT GGT GA 3'		665-684	
5) CgA sense: 5' GCT CCA AGA CCT CGC TCT CC 3'	583	316-335	Helman et al. (11)
6) CgA antisense: 5' GAC CGA CTC TCG CCT TTC CG 3'		878-897	

PCR, polymerase chain reaction.

in the Department of Pathology (University of Turin) specific for an N-terminal sequence of the sst2 (shared by both A and B receptor isoforms). The octapeptide EPYYDLTS, corresponding to amino acids 35 to 42 of the human receptor (and differing by one amino acid from the mouse sequence), was synthesized, having a lysin added to the N-terminal. This sequence was similar to that used by other groups to produce polyclonal antibodies (17,18,27). This sequence was rather short but made it possible to avoid extensive homology with sst1. In addition, according to a genbank search using FASTA (28), this protein sequence is unique to human sst2 and has a partial homology only with rat and human nuclear receptor retinoid orphan nuclear receptor-beta (a protein having nuclear localization). Three Balb/c mice were immunized with the peptide conjugated to keyhole limpets hemocyanin (KLH) (Sigma, St. Louis, MO) following the standard procedure. After the first intrasplenic injection (100  $\mu$ g of protein) at time 0, the mice were intraperitoneally injected six times with the peptide-KLH conjugate (150  $\mu$ g) in the presence of Freund adjuvant. The reactivity of the sera from each animal was evaluated using an enzyme-linked immunosorbent assay, using the peptide coated onto the plastic. The hybridomas were produced by somatic fusion of immunized splenocytes with the mouse myeloma cell line Ag8.X63.653, following the standard technique (21). The monoclonal antibodies of interest were selected on the basis of the reactivity with the target peptide and with appropriate tissue sections. The latter included formalin-fixed and paraffin-embedded sections of pituitary gland and pancreatic islets and were analyzed by means of immunoperoxidase staining. Parallel control experiments were also performed by staining serial sections of these tissues, omitting the primary antibody or with the preimmune serum or with the antibody preadsorbed with high concentrations (1 mg/mL) of the antigenic peptide. In addition, the selected monoclonal antibodies (coded 10C6 and 10G4), both of IgM isotype, were further characterized by Western blotting. Membranes were prepared from stable transfected Chinese hamster ovary (CHO)-K1 cells, individually expressing recombinant human somatostatin receptors (sst1 to sst5). Western blotting was performed as previously described (36). The monoclonal

antibody was used as culture supernatant at 1:3 dilution for 2 hours at room temperature in Tris-buffered saline (TBS), supplemented with 0.1% polysorbate 20. Blots were washed in TTBS and incubated with peroxidase-conjugated goat antimouse IgM, diluted 1:1,500 for 90 minutes at room temperature. Then, blots were washed in TTBS and immunocomplexes were visualized using ECL following manufacturer's instructions (Amersham, Bucks, UK).

A second polyclonal antibody was produced that had been characterized previously (35,36). This antibody (coded K230) was raised in sheep and was specific for a sequence of the C-terminal portion of the sst2A (KSRL-  
NETTETQRTLLNEDLQ, amino acids 347 to 366).

#### Immunohistochemistry

Sections 4 or 5  $\mu$  thick, adjacent to those used for conventional histopathologic examination and immunostaining for neuroendocrine markers, were collected onto poly-L-lysine-coated slides. The proliferative activity of the tumors was assessed by means of Ki67 immunostaining (clone MIB1, Immunotech, Marseille, France), diluted 1:10 after microwave-based antigen retrieval in citrate buffer. The ascitic fluid of monoclonal antibody 10G4 was used in this study and was applied to tissue sections with prior antigen retrieval (three 3-minute passages in a microwave oven at 800 W in citrate buffer pH 6.0), at the dilution of 1:10,000 or 1:12,000 for 30 minutes at room temperature. The antiserum coded K230 was applied overnight at a dilution of 1:300 with no prior antigen retrieval. The immune reactions were then revealed with the immunoperoxidase technique (13) using the streptavidin-peroxidase kit and diaminobenzidine as chromogen. A weak nuclear counterstain or no counterstain was used in parallel sections. Control stainings for both antibodies included immunoperoxidase of serial sections using preimmune serum or antibody preadsorbed with the antigen or buffer instead of the primary antibody.

#### In Situ Hybridization

Selected tumors (12 samples) were also analyzed for sst2 mRNA expression by means of a nonradioactive, tyramide deposition-based ISH technique. The proce-



cedure of amplification was modified from procedures reported by Kerstens et al. (16), Speel et al. (38), and the GenPoint (biotinyl-tyramide) manufacturer (Dako, Glostrup, Denmark). Briefly, 5- $\mu$ m-thick paraffin sections were collected onto silane-coated slides and deparaffinized through xylene and graded alcohols to phosphate buffer saline (PBS). The slides were then incubated for 5 minutes in a microwave oven at 800 W in citrate buffer pH 6.0. After washing in PBS, they were digested with proteinase K (1  $\mu$ g/mL) for 10 minutes at 23°C. Endogenous peroxidase activity was blocked with 3% hydrogen peroxide and endogenous biotin was blocked using avidin-blocking reagent for 15 minutes followed by washing in PBS and biotin-blocking reagent for 15 minutes (3). Sections were then prehybridized for 1 hour at room temperature in a mixture composed of 4 $\times$  SSC, 50% formamide, Denhardt's 1 $\times$ , dextrane sulfate 5 $\times$ , 500  $\mu$ g/mL salmon sperm DNA, and 250  $\mu$ g/mL tRNA. Hybridization took place overnight at 42°C in a solution containing the specific probe at a concentration of 1 pmol/mL. The probe was a digoxigenin-labeled 48-base oligonucleotide (32), complementary to positions 91 to 139 of the human *ssr2* gene (41). After hybridization, excess hybridization buffer and coverslips were removed by a rapid wash in 4 $\times$  SSC followed by stringent washing in 0.1 $\times$  SSC for 10 minutes at 42°C. The hybrids were revealed by the following incubation steps: peroxidase-labeled antidigoxigenin (diluted 1:100 in PBS) for 30 minutes at room temperature, biotinylated tyramide (diluted 1:5 in PBS) for 15 minutes at room temperature, and peroxidase-labeled streptavidin for 15 minutes at room temperature. Diaminobenzidine was used as chromogen. Controls for ISH included staining of serial sections with sense probe, an unrelated probe (EBER-1 of the Epstein-Barr virus), and omission of the probe in the hybridization mixture, with all other experimental conditions identical to the procedure described here.

## RESULTS

### Reverse Transcriptase Polymerase Chain Reaction

All neuroendocrine tumors, but no nonneuroendocrine lung carcinomas, were positive for CgA mRNA (Fig. 1). *Sst2* mRNA was amplified in 22 of 26 samples of neuroendocrine tumor. The signals had variable intensities (Fig. 2) and were weak in moderately or poorly differentiated tumors (mostly in SCCs). No amplification was obtained in no-template or no-reverse transcriptase experiments. Control samples (nonneuroendocrine lung carcinomas proven by negative CgA RT-PCR) were weakly positive for *ssr2* in 5 of 20 samples only (including 3 adenocarcinomas, 1 squamous, and the large-cell anaplastic carcinoma) (Fig. 3). These differences were statistically significant ( $P < 0.01$ ) by  $\chi^2$  test.

### Characterization of Monoclonal Antibodies to *ssr2*.

Several clones were identified having a positive binding by enzyme-linked immunosorbent assay and a parallel immunoreactivity on formalin-fixed paraffin-embedded human endocrine tissues (pituitary and pancreatic islets). In Western blotting experiments, two clones (coded 10C6 and 10G4) specifically developed a band at approximately 70 kD. When the antibodies were used against CHO-transfected cells expressing recombinant somatostatin receptors 1 through 5, a specific band corresponding to *ssr2* (at approximately 70 kD) was revealed by the monoclonal antibody 10G4. Monoclonal antibody 10C6 developed a strong band with *ssr2* but displayed a weaker reactivity also with *ssr1*, 3, and 5, at least in the present experimental conditions (Fig. 4 A,B). The same antibodies were also tested by means of immunoperoxidase staining on formalin-fixed, paraffin-embedded samples of normal human pituitary gland and pancreas. Monoclonal antibody 10G4 gave good results in immunohistochemistry and was used at increasing di-

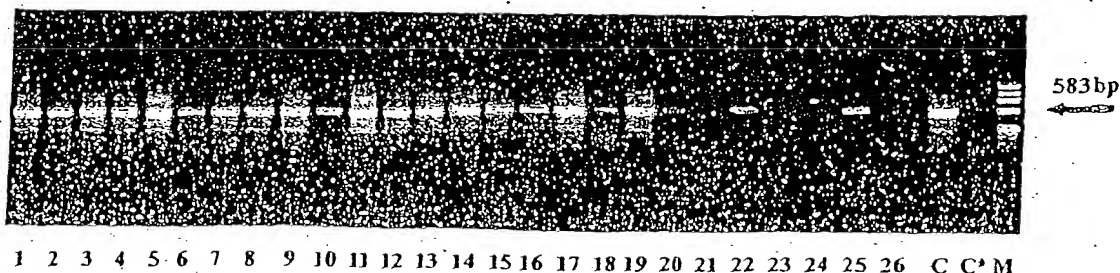


FIG. 1. Reverse transcriptase polymerase chain reaction for chromogranin A (CgA) mRNA in 26 samples of neuroendocrine tumor of the lung. Numbers in each lane correspond to sample numbers in Table 2. CgA mRNA is amplified at 429 bp. C and C\* stand for positive (neuroendocrine colon carcinoma cell line, H716) and negative (distilled water) controls, respectively. The last column to the right represents the molecular weight marker. All samples are positive with a variable intensity of the amplification band.

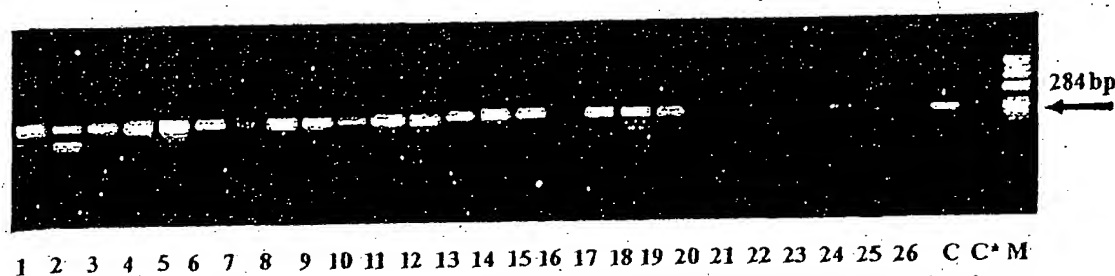


FIG. 2. Reverse transcriptase polymerase chain reaction for *sst2* mRNA in 26 samples of neuroendocrine tumor of the lung. Numbers in each lane correspond to sample numbers in Table 2. *sst2* mRNA is amplified at 284 bp. C and C\* stand for positive (a neuroblastoma) and negative (distilled water) controls, respectively. The last column to the right represents the molecular weight marker. Twenty-two of 26 samples are positive with a variable intensity of the amplification band.

lutions (up to 1:15,000) with specific staining. Using thin sections (approximately 4  $\mu$ m), a strong membrane-bound and peripheral cytoplasmic immunoreactivity was found in an adenohypophyseal cell population (corresponding to growth hormone-secreting cells, as confirmed by double immunohistochemical analyses) and in pancreatic islets (Fig. 4 C,D). In the latter, the staining was apparently not restricted to a specific hormone-producing cell type and had a peripheral cytoplasmic or membrane distribution. Exocrine pancreatic cells (both acinar and ductal) were only occasionally immunostained. Immunohistochemistry performed on serial control sections, either omitting the primary antibody or using the preimmune serum or antibodies preabsorbed with the synthetic peptide, was negative in both tissues. Monoclonal antibody 10C6 had a relatively higher background staining at similar dilutions.

#### Immunohistochemistry

The antibodies to *sst2* (monoclonal antibody 10G4 and polyclonal K230) gave slightly different immunoreactions in 25 samples, and staining was not done in 1 sample because of lack of residual paraffin blocks. The monoclonal antibody 10G4 stained 21 of 25 samples, the

negative samples being 1 atypical carcinoid and 3 SCCs (Fig. 5). The tumors had 5% to 25% of the neoplastic cells immunoreactive. The staining was at the periphery of the cytoplasm, and omitting the counterstain its membrane-bound distribution was better outlined in most samples (Fig. 6). One sample of atypical carcinoid (no. 21) was focally immunoreactive for *sst2*, despite negative RT-PCR findings. Conversely, sample no. 26 was immunohistochemistry negative and RT-PCR positive. The antiserum anti-*sst2A* (code K230) gave positive signal in 19 of 25 samples, in 5% to 60% of the neoplastic cell population (Fig. 7). The location of the staining was at the membrane level associated with a weak cytoplasmic reactivity. The same pattern was seen in positive controls, e.g., pancreatic islets (Fig. 7, inset). Two samples (nos. 19 and 26) were negative in spite of a positive RT-PCR signal. Two other tumors (nos. 21 and 22), apparently devoid of *sst2* mRNA, showed a small percentage of immunoreactive cells. Incidentally, one of these latter samples (no. 21) was also immunoreactive with monoclonal antibody 10G4 (Table 2).

The five control samples positive by RT-PCR were also reactive with the antibodies. The type of immunocytochemical location of *sst2* receptors was similar to that described here, being a peripheral cytoplasmic stain-

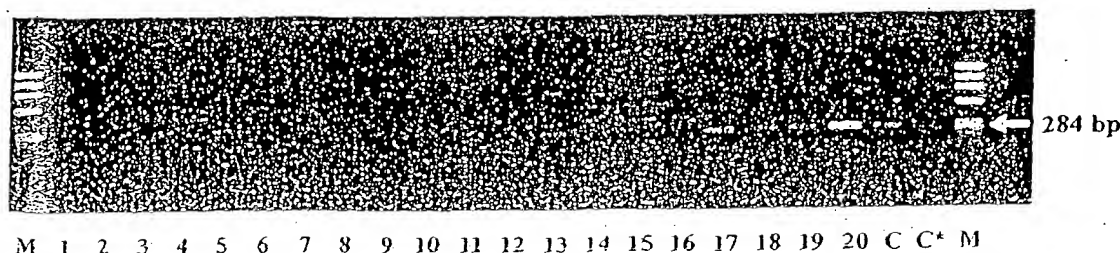


FIG. 3. Reverse transcriptase polymerase chain reaction for *sst2* mRNA in 20 control samples of nonneuroendocrine lung carcinoma. Five of 20 samples show a weak band at 284 bp corresponding to *sst2* mRNA. Control columns (C and C\*) are identical to those in Fig. 2.

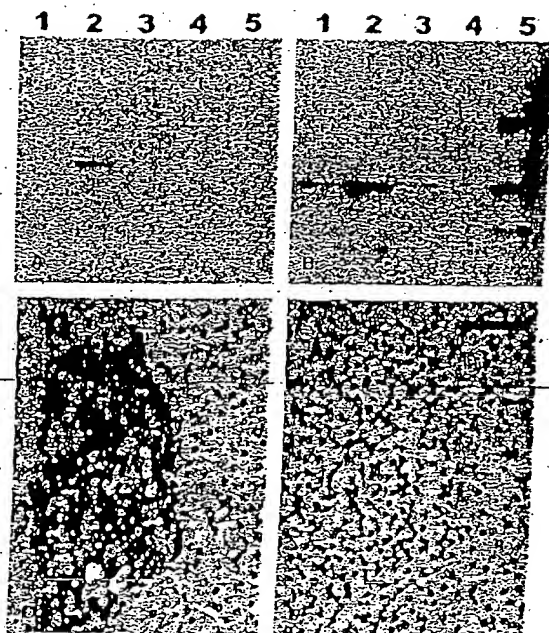


FIG. 4. Western blot analysis of monoclonal antibody clones 10G4 and 10C6 against sst2 in Chinese hamster ovary cells transfected with recombinant sst 1 through 5 (numbers of each column correspond to receptor type). Monoclonal antibody 10G4 shows a specific band at approximately 70 kD for sst2 only (A) as opposed to monoclonal antibody 10C6, which strongly reacts with sst2 but also has some degrees of cross-reactivity with sst 1, 3, and 5 (B). The lower figures show control formalin-fixed paraffin-embedded pancreatic islets immunostained with monoclonal antibody 10G4 without (C) and with (D) preadsorption with the peptide antigen, respectively. The majority of endocrine cells show a membrane-bound immunoreactivity (C) (immunoperoxidase). Bar: 90 µm.

ing present in 40% to 70% of neoplastic cells. A weak and focal staining was also observed in five of the remaining RT-PCR-negative samples, when the antibody K230 was used (but not when the monoclonal was employed).

Several cells in peritumoral tissues were occasionally stained. Ciliated cells of bronchial mucosa had a peripheral staining at the cilia border. Mucous glands were negative. Rare chondrocytes had a membrane staining. The wall of peritumoral as well as of occasional distant vessels was stained at the endothelium level and in occasional smooth muscle cells.

The reactivity of both antibodies was abolished in serial sections when the reagents were preabsorbed with the respective synthetic peptides, but not when an unrelated peptide was used. The peritumoral bronchial mucosa had a focal staining of ciliated cells with both antibodies. This reactivity disappeared when the preabsorbed antibody was applied.

#### In Situ Hybridization

Eight of 12 samples stained by ISH were positive for sst2 mRNA. The mRNA was present in a percentage of cells (ranging from 10% to 40%) and gave a weak signal (Fig. 8), despite the amplification provided by the tyramide-based procedure. The background level was minimal using diluted biotinylated tyramide. Control sections stained with sense probe or an unrelated probe, or omitting the probe, were consistently negative.

#### Clinical Data

Clinicopathologic data are summarized in Table 2. At follow-up, the majority of patients with typical carcinoids are free from disease 1 to 11 years after surgery. Two patients are alive with stable metastatic disease. Patients affected by atypical carcinoids had disease progression in one third of samples. Finally, patients with SCC had fatal outcomes within 1 year from diagnosis (except the recent sample). Eight patients had preoperative octreotide scintigraphy performed at the time of diagnosis. All patients had positive octreoscan findings, and, in these patients, also the tumor was positive by RT-PCR and immu-



FIG. 5. sample no. 25 (small cell carcinoma). Absence of immunoreactivity for sst2 with the monoclonal 10G4. This sample was also negative by reverse transcriptase polymerase chain reaction and in situ hybridization. (Immunoperoxidase in a formalin-fixed paraffin-embedded sample. Nuclei slightly counterstained with hemalum.) Bar: 45 µm.

nohistochemistry or ISH. In addition, three of these patients received octreotide therapy administered at the time of tumor recurrence or metastatic spread. Stable disease is recorded at follow-up more than 5 years after diagnosis.

### Correlations

Overall, complete overlapping (i.e., RT-PCR, ISH, and immunohistochemistry with two antibodies) between *sst2* gene and protein expression was obtained in 21 of 25 samples (84%) and between RT-PCR results and immunohistochemical findings with at least one of the antibodies in 24 of 25 samples (96%). The monoclonal antibody 10G4 looked highly sensitive, being able to stain all but one sample (no. 26) (95%) positive for *sst2* mRNA by RT-PCR. *Sst2* expression, at mRNA as well as at protein levels, was reduced in high grade tumors, with SCCs being weakly positive in only two of five samples. Decreasing expression of *sst2* appears to cor-

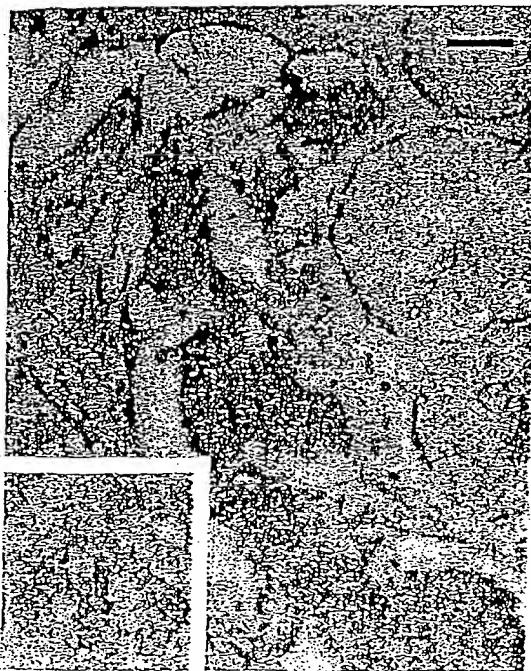


FIG. 6. sample no. 16 (typical carcinoid). Immunohistochemical detection of *sst2* by means of monoclonal antibody 10G4. The neoplastic cells have a peripheral cytoplasmic staining and membrane positivity in some cells, whereas the peribronchial gland adjacent to the tumor is unreactive. (Immunoperoxidase in a formalin-fixed paraffin-embedded sample. Nuclei slightly counterstained with hemalum.) Bar: 45  $\mu$ m. The membrane-bound distribution of the immunostaining is better outlined in a parallel section stained for monoclonal antibody 10G4 omitting nuclear counterstain (inset).



FIG. 7. Same sample as in Fig. 6. Immunohistochemical detection of *sst2* by means of the polyclonal antibody K230. The immunostaining is more intense at the cell border (arrows), as observed with the monoclonal antibody. In the inset, a pancreatic islet, used as positive control, shows a predominant membrane-bound immunostaining of many neuroendocrine cells. (Immunoperoxidase in a formalin-fixed paraffin-embedded sample. Nuclei slightly counterstained with hemalum.) Bar: 45  $\mu$ m.

relate with high tumor grade and elevated proliferative activity, but not with other parameters such sex, age, or tumor size.

### DISCUSSION

In this study, the presence of *sst2* mRNA has been demonstrated in a series of resected neuroendocrine tumors of the lung by means of RT-PCR and confirmed by a sensitive nonradioactive tyramide-based ISH procedure and by immunohistochemistry with anti-*sst2* antibodies. Samples of both carcinoid tumors and SCCs were *sst2* positive, although a reduced or absent signal was observed in poorly differentiated (small-cell) carcinomas. This is the first study of *sst2* expression in a relatively large series of neuroendocrine tumors of the lung. Single samples of human carcinoids and SCCs (including cell lines of the latter) had previously been analyzed and found to express *sst2* (7,15,30,32,33,39,42). Several methods have been used to detect these receptors and partially overlapping results were obtained.

In the present study, the expression of high amounts of

TABLE 2. Clinicopathologic data and somatostatin receptor type 2 (sst2) expression in 26 cases of neuroendocrine lung tumors

Patient no.	Diagnosis	Sex/age	Size (cm)	Follow-up (mo)	CgA IHC	CgA RT-PCR	SYP IHC	Ki67 IHC*	sst2 RT-PCR	sst2 IHC Mab(10G4)	sst2 IHC (K230 Ab)
1	WD NET	F/35	3.5	NED 90	+	+++	+ F	1.5	+++	+	+
2	WD NET	F/29	4	NED 45	+	++	+ F	3	++	+	+
3	WD NET	F/41	2.5	NED 70	+	++	+	0.1	+++	+	+ F
4	WD NET†	F/25	4	NED 21	+	++	+	2.6	+++	+ F	+ F
5	WD NET†	M/58	3.8	NED 23	+	+++	+	13	+++	+ F	+ F
6	WD NET	M/52	2.5	NED 42	+	++	+	4.5	+++	+ F	+ F
7	WD NET	F/69	3.5	NED 47	+	++	+	1	+	+	+
8	WD NET	M/29	3	NED 70	+	+++	+	1.5	+++	+	+
9	WD NET	M/27	4	NED 108	+	+++	+	NT	+++	+	+
10	WD NE Ca†	M/66	8	AWD 55	+	++	+ F	1.1	++	+ F	+
11	WD NET†	F/29	2	AWD 56	+	+++	+	2.5	+++	+	+
12	WD NET†	F/32	3	NED 26	+	++	+	4	+++	+	+
13	WD NE Ca	M/60	3	NED 133	+	+++	+	1	+++	NT	NT
14	WD NET	M/28	4	NED 130	+	+++	+	1	+++	+	+
15	WD NET†	M/41	1.3	AWD 53	+	+++	+	1.5	+++	+	+
16	WD NET†	F/31	1	NED 13	+	++	++	2.6	+	+	+
17	WD NET	F/53	4	NED 24	+	+++	+ F	4	+++	+ F	+
18	WD NE Ca	M/62	3	NED 6	-	++	-	13	+++	+	+ F
19	WD NE Ca	F/73	5	DOD 20	+	+++	+	3	++	+	-
20	SCC	M/57	6	DOD 12	+	++	+	45	+	+ F	+ F
21	SCC	M/51	4.5	DOD 5	+	+	+	35	-	+ F	+ F
22	WD NE Ca	M/60	6	NED 51	+	++	+	1.5	-	-	+
23	SCC	F/56	6	DOD 11	+	+	+	50	-	-	-
24	WD NE Ca	M/77	2.5	NED 21	+	+	+	24	+	+ F	+ F
25	SCC	M/57	5	DOD 10	+	++	++	80	-	-	-
26	SCC†	M/68	11	recent case	+	+	++	71	+	-	-

AWD, alive with disease; CgA, chromogranin A; DOD, died of disease; + F, focal: positive in <5% of cells; IHC, immunohistochemistry; Mab, monoclonal antibody; NECa, Neuroendocrine carcinoma; NED, no evidence of disease; NET, neuroendocrine tumor; NT, not tested; RT-PCR, reverse transcriptase polymerase chain reaction; SCC, Small-cell lung carcinoma; SYP, synaptophysin; WD, well differentiated.

\* Ki67 IHC: values correspond to percentage of positive nuclei of neoplastic cells.

† Patients who had preoperative octreoscan performed.

‡ Patients who had octreoscan performed and octreotide treatment.

sst2 mRNA was confirmed in well to moderately differentiated neuroendocrine tumors, in agreement with the results obtained by Reubi et al. (32) by means of radioactive ISH. The presence of sst2 mRNA in SCC had never been reported in human specimens, except for two samples included in Reubi et al.'s series (32). Although the data on cell lines support the observation that SCCs contain sst2 (42), slightly discrepant results were found in some of samples described here. Unfortunately, SCCs are rarely operated on, and therefore it is difficult to collect a large number of surgical specimens. The five samples studied in the current series by means of RT-PCR had a low amount (two samples) or absent (three samples) sst2 mRNA. This could be the result of the extensive necrosis commonly present in such tumor types. However, because care was taken to freeze fragments that were macroscopically devoid of necrotic areas, a more likely hypothesis is that sst2 expression is reduced in poorly differentiated tumors. Recently, Reisinger et al. (29) showed that the uptake of somatostatin analogues in patients with SCC undergoing chemotherapy is significantly lower, and therapeutic external factors may affect the receptor status of individual tumors. In addition, the uptake of somatostatin analogues

in metastatic deposits of SCC has been shown to be low or absent (29,2). The present findings suggest that the sst2 mRNA content is related to the degree of tumor differentiation. These data must be confirmed in larger series of nonneuroendocrine tumors to ascertain whether the observed loss or decrease of sst2 expression in neuroendocrine tumors is a common event linked to neoplastic dedifferentiation. In addition, further studies are needed to assess the functionality of such receptors, by comparing the profile of sst2 expression in tumor tissues with binding assays employing labeled somatostatin and with the clinical response to diagnostic and therapeutic administration of somatostatin analogues.

To this purpose, several investigators have demonstrated a correlation between clinical imaging or response to somatostatin analogue treatment and sst2 mRNA content in single samples of carcinoid tumors (15,22). Northern blotting and ISH were the techniques used for sst2 mRNA identification. This kind of correlation is useful for selecting patients for somatostatin analogue treatment, although the demonstration of receptor mRNA in a cell does not imply per se that the receptor is fully functional.

The present study relied on a highly sensitive tech-

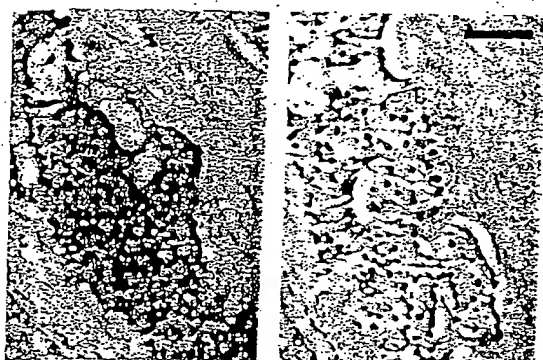


FIG. 8. sample no. 11 (typical carcinoid). In-situ-hybridization (ISH) for *sst2* mRNA shows a weak cytoplasmic staining (A) in most tumor cells. An ISH performed with an unrelated probe was negative in a serial section of the same tumor (B). This sample was strongly positive by reverse transcriptase polymerase chain reaction for *sst2* mRNA and by immunohistochemistry. (Nonradioactive ISH revealed by peroxidase and diaminobenzidine, as substrate. Nuclei counterstained with hemalum.) Bar: 75  $\mu$ m.

nique, RT-PCR, to identify all samples bearing even small amounts of *sst2* mRNA. Indeed, in a previous study, single samples exhibiting octreotide-binding sites had no demonstrable *sst2* mRNA by means of ISH, possibly due to the low sensitivity of the ISH procedure (34). The RT-PCR has shown *sst2* mRNA transcripts in the majority of samples here studied. Only four samples were negative, all belonging to poorly differentiated high grade tumors, which usually follow an aggressive course. A decrease of *sst2* mRNA expression in association with neuroendocrine tumor dedifferentiation had also been reported in neuroblastomas (37). In the above report, as well as in the current study, samples having an unfavorable prognosis were found to contain a relatively low amount of *sst2* mRNA, as compared with well-differentiated tumors.

In the current sample series, eight samples were investigated before surgery with radiolabeled octreotide. Despite the low figures, all the samples positive at the diagnostic procedure had a strong RT-PCR signal for *sst2* mRNA. Three of eight patients were also responsive to octreotide treatment administered at the time of relapse or metastatic spread. More extensive correlative clinicopathologic studies on the *sst* status are needed to better define the tissue distribution of somatostatin binding sites and their potential clinical role in the treatment of patients.

*Sst2* evaluation by means of ISH (14,32) or RT-PCR (26,37) is a highly sensitive and reliable procedure. Unfortunately, these techniques have limitations because frozen tissue is needed for some of them, and radioactive material or costly and time-consuming methods are nec-

essary for others. Immunohistochemical analysis of *sst2* by means of specific antibodies represents an ideal, cheap, and rapid alternative, easily applicable to archival material. For these reasons, several investigators have raised polyclonal antibodies specific for *sst* (8,10,15,17,18). In the current study, tested tumor fragments adjacent to those snap frozen for RT-PCR analysis were tested with a polyclonal antibody against a C-terminal portion of the *sst2A* splice variant (35,36). In addition, a monoclonal antibody was produced in the Department of Pathology (University of Turin) against an N-terminal sequence of the human *sst2*. This antibody was the first monoclonal developed against *sst2* and was shown to be highly specific for *sst2* in Western blot and immunohistochemical analysis. Both the monoclonal and the polyclonal antibodies specifically reacted with all samples also positive by RT-PCR (with minor discrepancies in two samples, likely due to tumor heterogeneity). The observed correlation between RT-PCR and immunohistochemistry indicates that the latter may be a reliable diagnostic tool and may allow immunohistochemical investigation for *sst2* even in small biopsy samples. This in turn may enable a rapid screening of *sst2*-positive tumors for medical treatment with somatostatin analogues.

Having confirmed in a relatively large series that the vast majority of neuroendocrine tumors of the lung contain variable amounts of *sst2* mRNA, a final comment is deserved for *sst2* expression in nonneuroendocrine lung carcinomas. No data have been reported thus far in the literature concerning normal human lung, although in the present study some bronchial cells of peritumoral parenchyma were positive for *sst2* when immunohistochemical analysis was performed with either antibody. The staining was specific because it was abolished using preabsorbed antibodies. Therefore, it is likely that normal human lung tissue contains *sst2*. This might be confirmed by alternative techniques (e.g., Western blot, RT-PCR). However, in situ morphologic procedures, such as those employed here, have definite advantages. In fact, the lung is rich in vessels, and in several tissues (either in tumoral or in inflammatory-reactive conditions) the vessels were recently shown to contain *sst* (6).

A low expression of *sst2* was found in 25% of lung carcinomas of nonneuroendocrine type investigated in the present study by means of RT-PCR. Therefore, *sst* type 2, at least, does not appear to be extensively expressed in nonneuroendocrine carcinomas of the lung. However, because two tumors in the control group (a squamous carcinoma and an adenocarcinoma, respectively) had positive octreotide scintigraphy, but no *sst2* mRNA, it is plausible that a heterogeneous distribution of *sst* occurs in nonneuroendocrine lung tumors. Other receptor types may be expressed in these tumors and may be responsible for the positive results in diagnostic testing. Because *sst5* is also known to bind somatostatin



analogues, such as octreotide, with high affinity (24), the expression of this receptor type will be investigated in future studies. □

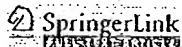
**Acknowledgments:** The authors are grateful to Dr. P.L. Filosso (Turin) for clinical data, to Mrs. M. Cerrato and Miss S. Solero for skilful technical help, and to Mr. A. Grua for the photographs.

## REFERENCES

- Abbona GC, Papotti M, Viberù L, Macrì L, Stella A, Bussolati G. Chromogranin A gene expression in non-small cell lung carcinomas. *J Pathol* 1998;186:1-6.
- Berenger N, Moretti JL, Boaziz C, Vigneron N, Morete JF, Breaux JL. Somatostatin receptor imaging in small cell lung cancer. *Eur J Cancer* 1996;32:1429-31.
- Bussolati G, Gugliotta P, Volante M, Pace M, Papotti M. Retrieved endogenous biotin: a novel marker and potential pitfall in diagnostic immunohistochemistry. *Histopathology* 1997;31:400-7.
- Capella C, Heitz PU, Hoster H, Solcia E, Kloppel G. Revised classification of neuroendocrine tumors of the lung, pancreas and gut. *Virchows Arch* 1995;425:547-60.
- Chingwin JM, Przybyla AE, Mac Donald RJ, Ruter W. Isolation of biologically active RNA from sources enriched in ribonuclease. *Biochemistry* 1979;18:5294-7.
- Denzler B, Reubi JC. Expression of somatostatin receptors in peritumoral veins of human tumors. *Cancer* 1999;85:189-98.
- Fujita T, Yamaji Y, Sato M, Miura K, Takahara J. Gene expression of somatostatin receptor subtypes, SST1 and SST2, in human lung cancer cell lines. *Life Sci* 1994;55:1797-806.
- Gu WZ, Schonbrunn A. Coupling specificity between somatostatin receptor sst2A and G proteins: isolation of the receptor-G protein complex with a receptor antibody. *Mol Endocrinol* 1997;11:527-37.
- Gussow D, Rein R, Ginjaar I, Hochstenbach F, Seemann G, Kottman A, Ploegh HL. The human beta 2-microglobulin gene: primary structure and definition of the transcriptional unit. *J Immunol* 1987;139:3132-8.
- Helboe L, Møller M, Norregaard L, Schiodt M, Sidsen CE. Development of selective antibodies against the human somatostatin receptor subtypes sst1-sst5. *Brain Res Mol Brain Res* 1997;49:82-8.
- Helman JJ, Ahn TG, Levine MA, et al. Molecular cloning and primary structure of human chromogranin A (secretory protein I) cDNA. *J Biol Chem* 1988;263:11559-63.
- Holland LJ, Liu Q, Van Koetsveld PM, et al. Immunohistochemical detection of somatostatin receptor subtypes sst1 and sst2A in human somatostatin receptor positive tumors. *J Clin Endocrinol Metab* 1999;84:775-80.
- Hsu SM, Raine L, Fanger H. Use of avidin-peroxidase complex (ABC) in immunoperoxidase techniques: comparison between ABC and unlabelled antibody (PAP) procedures. *J Histochem Cytochem* 1981;29:577-80.
- Janson ET, Gobl A, Kalkreuth KM, Oberg K. A comparison between the efficacy of somatostatin receptor scintigraphy and that of in situ hybridization for somatostatin receptor subtype 2 messenger RNA to predict therapeutic outcome in carcinoid patients. *Cancer Res* 1996;56:2561-5.
- Janson ET, Sridsberg M, Gobl A, Westlin JE, Oberg K. Determination of somatostatin receptors subtype 2 in carcinoid tumors by immunohistochemical investigation with somatostatin receptor subtype 2 antibodies. *Cancer Res* 1998;58:2375-8.
- Kerstens HJM, Poddighe PJ, Hanselaar AGJM. A novel in situ hybridization signal amplification method based on the deposition of biotinylated tyramine. *J Histochem Cytochem* 1995;43:347-52.
- Krisch B, Feindt J, Mendeim R. Immunoelectronmicroscopic analysis of the ligand-induced internalization of the somatostatin receptor subtype 2 in cultured human glioma cells. *J Histochem Cytochem* 1998;46:1233-42.
- Kumar U, Laird D, Srikant CB, Escher E, Patel YC. Expression of the five somatostatin receptors (SSTR1-5) subtypes in rat pituitary somatotrophs: quantitative analysis by double-label immunofluorescence confocal microscopy. *Endocrinology* 1997;138:4473-6.
- Kwok S, Higuchi R. Avoiding false positives with PCR. *Nature* 1989;339:237-8.
- Lamberts SWJ, Hofland LJ, Koetsveld PM, Reubi JC, Bruining HA, Bakker WH, Krenning EP. Parallel in vivo and in vitro detection of functional somatostatin receptors in human endocrine pancreatic tumors: consequences with regard to diagnosis, localization and therapy. *J Clin Endocrinol Metab* 1990;71:566-74.
- Malavasi F, Funaro A, Bellone G, et al. Functional and molecular characterization by the CB04 monoclonal antibody of a cell surface structure exerting C3-complement receptor activity. *J Clin Immunol* 1985;5:412-20.
- Nilsson O, Kolby L, Wängberg B, et al. Comparative studies on the expression of somatostatin receptor subtypes, outcome of octreotide scintigraphy and response to octreotide treatment in patients with carcinoid tumours. *Br J Cancer* 1998;77:632-7.
- Pagani A, Forni M, Tonini GP, Papotti M, Bussolati G. Expression of members of the chromogranin family in primary neuroblastomas. *Diagn Mol Pathol* 1992;1:16-24.
- Panetta R, Greenwood MT, Warszynska A, et al. Molecular cloning, functional characterization and chromosomal localization of a human somatostatin receptor (somatostatin receptor type 5) with preferential affinity for somatostatin-28. *Mol Pharmacol* 1994;45:417-27.
- Papotti M, Macrì L, Bussolati G, Reubi JC. Correlative study on neuroendocrine differentiation and presence of somatostatin receptors in breast carcinomas. *Int J Cancer* 1989;43:365-9.
- Papotti M, Macrì L, Pagani A, Aloï F, Bussolati G. Quantitation of somatostatin receptor type 2 in neuroendocrine (Merkel cell) carcinomas of the skin by competitive RT-PCR. *Endocr Pathol* 1999;10:1-10.
- Patel YC, Panetta R, Escher E, Greenwood M, Srikant CB. Expression of multiple somatostatin receptor genes in AIT-20 cells: evidence for a novel somatostatin-28 selective receptor subtype. *J Biol Chem* 1994;269:1506-9.
- Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A* 1988;85:2444-8.
- Reisinger I, Bohuslavski KH, Brenner W, et al. Somatostatin receptor scintigraphy in small-cell lung cancer: results of a multicenter study. *J Nucl Med* 1998;39:224-7.
- Reubi JC, Kappeler A, Waser B, Laissac J, Hipkin RW, Schonbrunn A. Immunohistochemical localization of somatostatin receptors sst2A in human tumors. *Am J Pathol* 1998;153:233-45.
- Reubi JC, Mawet R, von Werder K, Torhorst J, Klijn JGM, Lamberts SWJ. Somatostatin receptors in human endocrine tumors. *Cancer Res* 1987;47:551-8.
- Reubi JC, Schaefer JC, Waser B, Mengod G. Expression and localization of somatostatin receptors SST1, SST2, SST3 messenger RNA in primary human tumors using in situ hybridisation. *Cancer Res* 1994;54:3455-9.
- Reubi JC, Waser B, Sheppard M, Macaulay V. Somatostatin receptors are present in small-cell but not in non-small-cell primary lung carcinomas: relationship to EGF receptors. *Int J Cancer* 1990;45:269-74.
- Schaefer JC, Waser B, Mengod G, Reubi JC. Somatostatin receptors subtypes sst1, sst2, sst3 and sst5 expression in human pituitary, gastroentero-pancreatic and neanary tumors: comparison of mRNA analysis with receptor autoradiography. *Int J Cancer* 1997;70:530-7.
- Schindler M, Holloway S, Humphrey PPA, Waldvogel H, Faull RL, Berger W, Emson PC. Localization of the somatostatin sst2A receptor in human cerebral cortex, hippocampus and cerebellum. *Neuroreport* 1998;9:521-5.
- Schindler M, Sellers LA, Humphrey PPA, Emson PC. Immunohistochemical localization of the somatostatin sst2(A) receptor in the rat brain and spinal cord. *Neuroscience* 1997;76:225-40.
- Sestini R, Orlando C, Peri A, et al. Quantification of somatostatin

- receptor type-2 gene expression in neuroblastoma cell lines and primary tumors using competitive reverse transcription-polymerase chain reaction. *Clin Cancer Res* 1996;2:1757-65.
38. Speil EDM, Saremaslani P, Roth J, Hopman AHN, Komminoth P. Improved mRNA in situ hybridization on formaldehyde-fixed and paraffin-embedded tissue using signal amplification with different haptenized tyramides. *Histochem Cell Biol* 1998;420:1-7.
39. Taylor JE, Theveniau MA, Bashirzadeh R, Reisine T, Eden PA. Detection of somatostatin receptor subtype 2 (SSTR2) in established tumors and tumor cell lines: evidence for sst2 heterogeneity. *Peptides* 1994;15:1229-36.
40. Travis WD, Gal AA, Colby TV, Klimstra DS, Falk R, Koss MN. Reproducibility of neuroendocrine lung tumor classification. *Hum Pathol* 1998;29:272-9.
41. Yamada Y, Post SR, Wang K, Tager H, Bell GI, Scino S. Cloning and functional characterization of a family of human and mouse somatostatin receptors expressed in brain, gastrointestinal tract and kidney. *Biochemistry* 1992;89:251-5.
42. Zhang CY, Yokogoshi Y, Yoshimoto K, Fujimaki Y, Matsumoto K, Saito S. Point mutation of the somatostatin receptor 2 gene in the human small cell lung cancer cell line COR-L103. *Biochem Biophys Res Commun* 1995;210:805-15.





**Expression of somatostatin receptor types 1-5 in 81 cases of gastrointestinal and pancreatic endocrine tumors. A correlative immunohistochemical and reverse-transcriptase polymerase chain reaction analysis.**

Papotti M, Bongiovanni M, Volante M, Allia E, Landolfi S, Helboe L, Schindler M, Cole SL, Bussolati G.

Department of Biomedical Sciences and Oncology, University of Turin, Via Santena 7, 10126 Turin, Italy. mauro.papotti@uni-to.it

Somatostatin receptors (SSTRs) have been extensively mapped in human tumors by means of autoradiography, reverse-transcriptase polymerase chain reaction (RT-PCR), in situ hybridization (ISH) and immunohistochemistry (IHC). We analyzed the SSTR type 1-5 expression by means of RT-PCR and/or IHC in a series of 81 functioning and non-functioning gastroenteropancreatic (GEP) endocrine tumors and related normal tissues. Moreover, we compared the results with clinical, pathological and hormonal features. Forty-six cases (13 intestinal and 33 pancreatic) were studied for SSTR 1-5 expression using RT-PCR, IHC with antibodies to SSTR types 2, 3, 5 and ISH for SSTR2 mRNA. The vast majority of tumors expressed SSTR types 1, 2, 3 and 5, while SSTR4 was detected in a small minority. Due to the good correlation between RT-PCR and IHC data on SSTR types 2, 3, and 5, thirty-five additional GEP endocrine tumors were studied with IHC alone. Pancreatic insulinomas had an heterogeneous SSTR expression, while 100% of somatostatinomas expressed SSTR5 and 100% gastrinomas and glucagonomas expressed SSTR2. Pre-operative biopsy material showed an overlapping immunoreactivity with that of surgical specimens, suggesting that the SSTR status can be detected in the diagnostic work-up. It is concluded that SSTRs 1-5 are heterogeneously expressed in GEP endocrine tumors and that IHC is a reliable tool to detect SSTR types 2, 3 and 5 in surgical and biopsy specimens.

PMID: 12021920 [PubMed - indexed for MEDLINE]



**P-cadherin overexpression is an indicator of clinical outcome in invasive breast carcinomas and is associated with CDH3 promoter hypomethylation.**

Paredes J, Albergaria A, Oliveira JT, Jeronimo C, Milanezi E, Schmitt FC.

Institute of Pathology and Molecular Immunology of Porto University (IPATIMUP), Braga, Portugal. [jparedes@ipatimup.pt](mailto:jparedes@ipatimup.pt)

**PURPOSE:** P-cadherin overexpression has been reported in breast carcinomas, where it was associated with proliferative high-grade histological tumors. This study aimed to analyze P-cadherin expression in invasive breast cancer and to correlate it with tumor markers, pathologic features, and patient survival. Another purpose was to evaluate the P-cadherin promoter methylation pattern as the molecular mechanism underlying this gene regulation. **EXPERIMENTAL DESIGN:** Using a series of invasive breast carcinomas, P-cadherin expression was evaluated and correlated with histologic grade, estrogen receptor, MIB-1, and p53 and c-erbB-2 expression. In order to assess whether P-cadherin expression was associated with changes in CDH3 promoter methylation, we studied the methylation status of a gene 5'-flanking region in these same carcinomas. This analysis was also done for normal tissue and for a breast cancer cell line treated with a demethylating agent. **RESULTS:** P-cadherin expression showed a strong correlation with high histologic grade, increased proliferation, c-erbB-2 and p53 expression, lack of estrogen receptor, and poor patient survival. This overexpression can be regulated by gene promoter methylation because the 5-Aza-2'-deoxycytidine treatment of MCF-7/AZ cells increased P-cadherin mRNA and protein levels. Additionally, we found that 71% of P-cadherin-negative cases showed promoter methylation, whereas 65% of positive ones were unmethylated ( $P = 0.005$ ). The normal P-cadherin-negative breast epithelial cells showed consistent CDH3 promoter methylation. **CONCLUSIONS:** P-cadherin expression was strongly associated with tumor aggressiveness, being a good indicator of clinical outcome. Moreover, the aberrant expression of P-cadherin in breast cancer might be regulated by gene promoter hypomethylation.

PMID: 16115928 [PubMed - in process]



**Mammary-derived growth inhibitor protein and messenger ribonucleic acid concentrations in different physiological states of the gland.**

Politis I, Gorewit RC, Muller T, Grosse R.

Department of Animal Science, Cornell University, Ithaca 14853.

Expression of mammary-derived growth inhibitor in tissue from lactating and involuting bovine mammary glands was investigated. Seventeen lactating, pregnant (220 to 272 d in gestation) cows were divided in two groups of 8 and 9 cows each. Cows of the first group were slaughtered while in lactation. Cows of the second group (9 involuting cows) were slaughtered at 13 to 52 d following sudden cessation of milking. High concentrations of mammary-derived growth inhibitor (.63% of the total protein) were detected in mammary tissue of lactating cows. Mammary-derived growth inhibitor (less than .10% of the total protein) was dramatically reduced during most of the involution period (13 to 45 d following cessation of milking). Mammary-derived growth inhibitor was again detected (.28% of the total protein) during the last stage of the involution (46 to 53 d after cessation of milking), which coincided with colostrum formation. When steady state concentrations of mammary-derived growth inhibitor mRNA were examined, the results obtained mirrored those obtained at the protein concentration. These data suggest that regulation of mammary-derived growth inhibitor occurs via modulation of the steady state concentration of its mRNA. Furthermore, there is a strong correlation between mammary-derived growth inhibitor expression and lactation in dairy cows.

PMID: 1500548 [PubMed - indexed for MEDLINE]

Comment in:

- J Invest Dermatol. 1994 Nov;103(5):742-4.

**T-cell receptor V beta-family usage in primary cutaneous and primary nodal T-cell non-Hodgkin's lymphomas.**

Preesman AH, Hu HZ, Tilanus MG, de Geus B, Schuurman HJ, Reitsma R, van Wichen DF, van Vloten WA, de Weger RA.

Department of Pathology, University Hospital Utrecht, The Netherlands.

To evaluate whether the expression of T-cell receptor (TCR) V beta families in eight cases of malignant T-cell lymphomas took place in a preferential manner, we analyzed four cases of mycosis fungoides (MF), the most common form of primary cutaneous T-cell non-Hodgkin's lymphomas (NHL), and four cases of primary nodal T-cell NHL. The usage of V beta families in T-cell populations was investigated on mRNA that was transcribed to cDNA using a C beta primer and reverse transcriptase. Subsequently, the specific usage of the families was analyzed by polymerase chain reaction (PCR) using combinations of the selected C beta-oligonucleotide primer and one of the family-specific V beta primers. Peripheral blood lymphocytes from four healthy volunteers and 1 "reactive" lymph node served as a control and expressed all 20 V beta families tested for. In T-cell lines, with restricted V beta expression, and in three patients with advanced MF, only one or two V beta families were expressed at the mRNA level. In an early MF lesion this monoclonal expression was absent: several V beta families were expressed with a weak intensity. This may indicate either a polyclonal origin of MF, or that too few monoclonal neoplastic cells were present in the tissue specimen. In the four nodal T-cell NHL, only one family could be clearly distinguished, whereas some of the other V beta families showed only a weak expression. These latter families represent the reactive T-cell component in the nodal T-cell NHL. Both in nodal T-cell NHL and in MF there was no preferential expression of a particular V beta family. There was a good correlation between PCR data and the expression of V beta-family protein products observed by immunohistochemistry on tissue sections of the T-cell lymphomas. All T-cell lines, three cases of MF, and three cases of nodal T-cell NHL showed a rearrangement of the TCR beta chain on DNA level.

PMID: 1331246 [PubMed - indexed for MEDLINE]



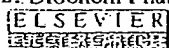
**Matrilin-3 in human articular cartilage: increased expression in osteoarthritis.**

Pullig O, Weseloh G, Klatt AR, Wagener R, Swoboda B.

Division of Orthopaedic Rheumatology, Department of Orthopaedics, University of Erlangen-Nuremberg, Rathsberger Str. 57, D-91054 Erlangen, Germany.  
oliver.pullig@med.uni-erlangen.de

**OBJECTIVE:** Matrilin-3 is a member of the recently described matrilin family of extracellular matrix proteins containing von Willebrand factor A-like domains. The matrilin-3 subunit can form homo-tetramers as well as hetero-oligomers together with subunits of matrilin-1 (cartilage matrix protein). It has a restricted tissue distribution and is strongly expressed in growing skeletal tissues. Detailed information on expression and distribution of extracellular matrix proteins is important to understand cartilage function in health and in disease like osteoarthritis (OA). **METHODS:** Normal and osteoarthritic cartilage were systematically analysed for matrilin-3 expression, using immunohistochemistry, Western blot analysis, in situ hybridization, and quantitative PCR. **RESULTS:** Our results indicate that matrilin-3 is a mandatory component of mature articular cartilage with its expression being restricted to chondrocytes from the tangential zone and the upper middle cartilage zone. Osteoarthritic cartilage samples with only moderate morphological osteoarthritic degenerations have elevated levels of matrilin-3 mRNA. In parallel, we found an increased deposition of matrilin-3 protein in the cartilage matrix. Matrilin-3 staining was diffusely distributed in the cartilage matrix, with no cellular staining being detectable. In cartilage samples with minor osteoarthritic lesions, matrilin-3 deposition was restricted to the middle zone and to the upper deep zone. A strong correlation was found between enhanced matrilin-3 gene and protein expression and the extent of tissue damage. Sections with severe osteoarthritic degeneration showed the highest amount of matrilin-3 mRNA, strong signals in in situ hybridization, and prominent protein deposition in the middle and deep cartilage zone. **CONCLUSION:** We conclude that matrilin-3 is an integral component of human articular cartilage matrix and that the enhanced expression of matrilin-3 in OA may be a cellular response to the modified microenvironment in the disease. Copyright 2002 OsteoArthritis Research Society International.

PMID: 11950247 [PubMed - indexed for MEDLINE]



**Up-regulation of mitochondrial peripheral benzodiazepine receptor expression by tumor necrosis factor alpha in testicular leydig cells. Possible involvement in cell survival.**

**Rey C, Mauduit C, Naureils O, Benahmed M, Louisot P, Gasnier F.**

INSERM U. 189, Faculte de Medecine Lyon-Sud, BP12, 69921 cedex, Oullins, France.

Porcine Leydig cells in primary cultures are resistant to tumor necrosis factor alpha (TNFalpha) cytotoxicity. Here we report that these cells can be rendered sensitive to TNFalpha killing by treatment with the translational inhibitor cycloheximide, suggesting the existence of proteins that can suppress the death stimulus induced by the cytokine. In search of these cytoprotective proteins, we focused on the constituents of the mitochondrial permeability transition pore (PT pore), whose opening has been shown to play a critical role in the TNFalpha-mediated death pathway. We found that TNFalpha up-regulated mRNA and protein expression of the mitochondrial peripheral benzodiazepine receptor (PBR), an outer membrane-derived constituent of the pore. A strong correlation was established between the resistance of the cells to TNFalpha killing and the density of PBR-binding sites. Concomitantly, TNFalpha down-regulated Bcl-2 mRNA and protein expression. As Bcl-2 has been shown to be an endogenous inhibitor of the PT pore, we hypothesize that the TNFalpha-induced up-regulation of PBR expression may compensate for the decrease in Bcl-2 levels to prevent the opening of the PT pore.

PMID: 11077046 [PubMed - indexed for MEDLINE]

**MetaPress**

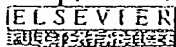
**GLUT1 messenger RNA and protein induction relates to the malignant transformation of cervical cancer.**

Rudlowski C, Becker AJ, Schroder W, Rath W, Buttner R, Moser M.

Dept of Gynecology and Obstetrics, University Hospital Heidelberg, Vossstr 7-9, D-69115 Heidelberg, Germany.

We studied whether induction of glucose transporters (GLUTs) 1 to 4 correlates with human papillomavirus (HPV)-dependent malignant transformation of cervical epithelium. Tissue samples of cervical intraepithelial neoplasia (CIN; grades 1 to 3), invasive carcinomas, and lymph node metastasis were examined. HPV typing was performed. Tissue sections were immunostained with GLUT1 to GLUT4 antibodies. Messenger RNA (mRNA) in situ hybridization confirmed GLUT1 protein expression. Weak expression of GLUT1 was found in nondysplastic HPV-positive and HPV-negative epithelium; significant expression was observed in preneoplastic lesions, correlating with the degree of dysplasia. In CIN 3 high-risk HPV lesions, cervical cancer, and metastasis, GLUT1 was expressed at highest levels with a strong correlation of GLUT1 mRNA and protein expression. Immunostains for GLUT2 to GLUT4 were negative. Cervical tumor cells respond to enhanced glucose utilization by up-regulation of GLUT1. The strong induction of GLUT1 mRNA and protein in HPV-positive CIN 3 lesions suggests GLUT1 overexpression as an early event in cervical neoplasia. GLUT1 is potentially relevant as a diagnostic tool and glucose metabolism as a therapeutic target in cervical cancer.

PMID: 14608894 [PubMed - indexed for MEDLINE]



**Expression and distribution of laminin alpha1 and alpha2 chains in embryonic and adult mouse tissues: an immunochemical approach.**

Sasaki T, Giltay R, Talts U, Timpl R, Talts JF.

Max-Planck-Institute for Biochemistry, Martinsried, D-82152, Germany..

Protein levels, mRNA expression, and localization of laminin alpha1 and alpha2 chains in development and in adult mice were examined. Recombinant fragments were used to obtain high-titer-specific polyclonal antibodies for establishing quantitative radioimmuno-inhibition assays. This often demonstrated an abundance of alpha2 chain, but also distinct amounts of alpha1 chain for adult tissues. The highest amounts of alpha1 were found in placenta, kidney, testis, and liver and exceeded those of alpha2. All other tissue extracts showed a higher content of alpha2, which was particularly high in heart and muscle when compared to alpha1. Content of gamma1 chain, shared by most laminins, was also analyzed. This demonstrated gamma1 chain levels being equal to or moderately exceeding the sum of alpha1 and alpha2 chains, indicating that these isoforms represent the major known laminin isoforms in most adult mouse tissues so far examined. Moreover, we found good correlation between radioimmuno-inhibition data and mRNA levels of adult tissues as measured by quantitative real-time reverse transcriptase-PCR. Embryonic tissues were also analyzed by radioimmuno-inhibition assays. This demonstrated for day 11 embryos comparable amounts of alpha1 and gamma1 and a more than 25-fold lower content of alpha2. This content increased to about 10% of alpha1 in day 13 embryos. The day 18 embryo showed in heart, kidney, and liver, but not yet in brain and lung; alpha1/alpha2 chain ratios comparable to those in adult tissues. Immunostaining demonstrated alpha1 in Reichert's membrane (day 7.5), while alpha2 could not be detected before day 11.5. These data were compared with immunohistochemical localization results on several more embryonic and adult tissue sections. Our results regarding localization are consistent with those of earlier work with some notable exceptions. This was in part due to epitope masking for monoclonal antibodies commonly used in previous studies in esophagus, intestine, stomach, liver, kidney, and spleen.

PMID: 11969289 [PubMed - indexed for MEDLINE]



FREE full-text article at  
[www.jbc.org](http://www.jbc.org)

**Discordant regulation of granzyme H and granzyme B expression in human lymphocytes.**

Sedelies KA, Sayers TJ, Edwards KM, Chen W, Pellicci DG, Godfrey DI, Trapani JA.

Cancer Immunology Laboratory, Peter MacCallum Cancer Centre, Locked Bag 1, A'Beckett Street, East Melbourne, 8006, Australia.

We analyzed the expression of granzyme H in human blood leukocytes, using a novel monoclonal antibody raised against recombinant granzyme H. 33-kDa granzyme H was easily detected in unfractionated peripheral blood mononuclear cells, due to its high constitutive expression in CD3(-)CD56(+) natural killer (NK) cells, whereas granzyme B was less abundant. The NK lymphoma cell lines, YT and Lopez, also expressed high granzyme H levels. Unstimulated CD4(+) and particularly CD8(+) T cells expressed far lower levels of granzyme H than NK cells, and various agents that classically induce T cell activation, proliferation, and enhanced granzyme B expression failed to induce granzyme H expression in T cells. Also, granzyme H was not detected in NK T cells, monocytes, or neutrophils. There was a good correlation between mRNA and protein expression in cells that synthesize both granzymes B and H, suggesting that gzmH gene transcription is regulated similarly to gzmB. Overall, our data indicate that although the gzmB and gzmH genes are tightly linked, expression of the proteins is quite discordant in T and NK cells. The finding that granzyme H is frequently more abundant than granzyme B in NK cells is consistent with a role for granzyme H in complementing the pro-apoptotic function of granzyme B in human NK cells.

PMID: 15069086 [PubMed - indexed for MEDLINE]



**BCL2 protein expression parallels its mRNA level in normal and malignant B cells.**

**Shen Y, Iqbal J, Huang JZ, Zhou G, Chan WC.**

Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, USA.

The regulation of B-cell lymphoma 2 (BCL2) protein expression in germinal center (GC) B cells has been controversial. Previous reports have indicated posttranscriptional regulation plays a dominant role. However, a number of recent studies contradicted these reports. Using real-time polymerase chain reaction (PCR) and Standardized Reverse Transcriptase-PCR (StaRT-PCR), we measured the level of mRNA expression in GC, mantle zone (MNZ), and marginal zone (MGZ) cells from laser capture microdissection. Both quantitative RT-PCR measurements of microdissected GC cells from tonsils showed that GC cells had low expression of BCL2 transcripts commensurate with the low protein expression level. These results are in agreement with microarray studies on fluorescence-activated cell sorter (FACS)-sorted cells and microdissected GC cells. We also examined BCL2 mRNA and protein expression on a series of 30 cases of diffuse large B-cell lymphoma (DLBCL) and found, in general, a good correlation. The results suggested that BCL2 protein expression is regulated at the transcriptional level in normal B cells and in the neoplastic cells in most B-cell lymphoproliferative disorders.

PMID: 15242877 [PubMed - indexed for MEDLINE]

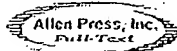
**Quantitative determinations of the steady state transcript levels of hexokinase isozymes and glucose transporter isoforms in normal rat tissues and the malignant tumor cell line AH130.**

Shinohara Y, Yamamoto K, Inoo K, Yamazaki N, Terada H.

Faculty of Pharmaceutical Sciences, University of Tokushima, Japan.  
yasuo@ph.tokushima-u.ac.jp

The steady state transcript levels of the four hexokinase (HK) isozymes and four glucose transporter (GLUT) isoforms were determined quantitatively by Northern analysis of RNA samples from rat tissues using synthetic fragments of the RNAs encoding the HK isozymes and GLUT isoforms. Results showed that the levels of HK isozyme transcripts were low in rat tissues, the level of that most highly expressed, the type I isozyme (HKI), in the brain being 0.025% of the total poly(A)+ RNA. A good correlation was found between the reported HK activities and the total amounts of transcripts encoding all HK isozymes in various tissues, showing that the HK activities in tissues can be estimated from the total amount of transcripts encoding HK isozymes. The proposed associated expressions of HK isozymes and GLUT isoforms in particular tissues were confirmed at their transcript levels. The steady state transcript levels of type II HK and the type I GLUT isoform in the malignant tumor cell line AH130 were also determined quantitatively.

PMID: 9459591 [PubMed - indexed for MEDLINE]



**UVA irradiation-induced activation of activator protein-1 is correlated with induced expression of AP-1 family members in the human keratinocyte cell line HaCaT.**

**Silvers AL, Bowden GT.**

Department of Radiation Oncology, Arizona Cancer Center, The University of Arizona, Tucson 85724, USA.

To determine whether the transcription factor activator protein-1 (AP-1) could be modulated by ultraviolet A (UVA) exposure, we examined AP-1 DNA-binding activity and transactivation after exposure to UVA in the human immortalized keratinocyte cell line HaCaT. Maximal AP-1 transactivation was observed with 250 kJ/m<sup>2</sup> UVA between 3 and 4 h after irradiation. DNA binding of AP-1 to the target 12-O-tetradecanoylphorbol-13-acetate response element sequence was maximally induced 1-3 h after irradiation. Both de novo transcription and translation contributed to the UVA-induced AP-1 DNA binding. c-Fos was implicated as a primary component of the AP-1 DNA-binding complex. Other components of the complex included Fra-2, c-Jun, JunB and JunD. UVA irradiation induced protein expression of c-Fos, c-Jun, Fra-1 and Fra-2. Phosphorylated forms of these induced proteins were determined at specific time points. A strong correlation existed between UVA-induced AP-1 activity and accumulation of c-Fos, c-Jun and Fra-1 proteins. UVA irradiation also induced c-fos and c-jun mRNA expression and transcriptional activation of the c-fos gene promoter. These results demonstrate that UVA irradiation activates AP-1 and that c-fos induction may play a critical role in the response of these human keratinocytes to UVA irradiation.

PMID: 11950097 [PubMed - indexed for MEDLINE]



# **Rat kidney glutamyl aminopeptidase (aminopeptidase A): molecular identity and cellular localization.**

**Song L, Ye M, Troyanovskaya M, Wilk E, Wilk S, Healy DP.**

Department of Pharmacology, Mount Sinai School of Medicine, City University of New York, New York 10029.

Glutamyl aminopeptidase [aminopeptidase A (EAP), EC 3.4.11.7] is an ectoenzyme that selectively hydrolyzes acidic amino acid residues from the amino terminus of oligopeptides. EAP activity is highest within the kidney and small intestine. The murine pre-B cell BP-1/6C3 and the human kidney glycoprotein gp160 differentiation antigens have been reported to have biochemical properties indistinguishable from EAP. It is not known, however, if rat kidney EAP is a homologue of these antigens or molecularly distinct. Using the reverse transcription-polymerase chain reaction method with oligonucleotide primers based on the BP-1/6C3 nucleotide sequence, we isolated a 450-bp partial cDNA from rat kidney poly(A)+ RNA. The partial cDNA encoded a predicted protein that was 92% and 86% identical to the murine BP-1/6C3 and human gp160 antigens, respectively; the amino acid sequence within the zinc-binding domain was completely conserved. Purification of EAP from rat kidney and microsequence analysis of a tryptic digest peptide fragment (18-mer) indicated that the fragment was highly similar to a region within the BP-1/6C3 and gp160 proteins. Northern blot hybridization and immunoblot analyses were also consistent with labeling of products the same size as reported for the BP-1/6C3 and gp160 antigens. There was a good correlation between the cellular distribution of EAP mRNA and EAP immunoreactivity, with proximal tubules and glomerular mesangial cells having the highest densities. These results indicate that rat kidney EAP is a species homologue of the murine BP-1/6C3 and human gp160 antigens. Furthermore, on the basis of its cellular localization, rat kidney EAP is likely to be involved in degradation of oligopeptides within the glomerulus and the glomerular filtrate. Since cells that express EAP also express receptors for angiotensin II, an intrarenal vasoactive hormone that is a substrate for EAP, these results further suggest that EAP may play a role in modulating the activity of intrarenal angiotensin II.

PMID: 7943354 [PubMed - indexed for MEDLINE]

**Tumor necrosis factor-alpha upregulates the prostaglandin E2 EP1 receptor subtype and the cyclooxygenase-2 isoform in cultured amnion WISH cells.**

Spaziani EP, Benoit RR, Tsibris JC, Gould SF, O'Brien WF.

University of South Florida Health Science Center, Department of Obstetrics & Gynecology, Tampa 33612, USA. [espazian@com1.med.usf.edu](mailto:espazian@com1.med.usf.edu)

Recent studies have demonstrated a strong correlation between infection and preterm labor. Preterm delivery is also associated with high levels of cytokines and prostaglandins in amniotic fluid. The purpose of this study was to investigate the effect of tumor necrosis factor-alpha (TNF-alpha) on the levels of cyclooxygenase, prostaglandin E2 production (PGE2), and expression of the PGE2 receptor subtype EP1 in amnion WISH cell culture. Amnion WISH cell cultures were incubated in increasing concentrations of TNF-alpha (0-50 ng/ml). Changes in cyclooxygenase and EP1 receptor proteins were evaluated by Western blot analysis. Changes in EP1 mRNA were evaluated by Northern blot, and culture fluid concentrations of PGE2 were estimated by enzyme immunoassay (EIA). EP1 protein ( $p<0.01$ ), EP1 mRNA ( $p<0.05$ ), cyclooxygenase-2 (COX-2) protein ( $p<0.001$ ), and PGE2 concentrations ( $p<0.01$ ) all increased with increasing concentrations of TNF-alpha. Changes in COX-1 protein were not observed following TNF-alpha-incubation. The results suggest that TNF-alpha may play a role in infection-induced preterm labor by its pleiotropic ability to simultaneously stimulate COX-2 activity, PGE2 concentrations, and PGE2 EP1 receptor levels in human amnion.

PMID: 9877447 [PubMed - indexed for MEDLINE]



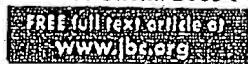
**Transcriptional activity of potent glucocorticoids: relevance of glucocorticoid receptor isoforms and drug metabolites.**

Spika I, Hammer S, Kleuser B, Korting HC, Schafer-Korting M.

Institut für Pharmazie, Abteilung für Pharmakologie und Toxikologie, Freie Universität Berlin, Berlin, Germany.

As compared to standard glucocorticoids (GC), prednicarbate (PC) is favorable in the treatment of eczema due to its high benefit/risk ratio. The remarkable anti-inflammatory effects of PC are in strong contrast to its reported low glucocorticoid receptor (GR) binding affinity. In transfected COS-7 cells we related the transcriptional potencies of PC, its metabolites and conventional GC to their receptor binding properties. Moreover, the expression pattern of the human GR isoform hGRalpha and its mutual dominant negative inhibitor hGRbeta in skin cells have been investigated as well as the influence of hGRbeta on receptor binding and transactivation. hGRalpha mRNA and protein was largely overexpressed in skin cells. hGRbeta showed no influence on hGRalpha binding and transactivation. Concentration response curves indicated the greater transactivation potency of betamethasone 17-valerate followed by dexamethasone and prednisolone 17-ethylcarbonate. Native PC appeared almost as potent as dexamethasone. With both a strong correlation was observed between transactivation and GR binding. Copyright 2003 S. Karger AG, Basel

PMID: 12677094 [PubMed - indexed for MEDLINE]



**Specific inhibition of AQP1 water channels in isolated rat intrahepatic bile duct units by small interfering RNAs.**

Splinter PL, Masyuk AI, LaRusso NF.

Center for Basic Research in Digestive Diseases, Division of Gastroenterology and Hepatology, Mayo Medical School, Clinic, and Foundation, Rochester, Minnesota 55905, USA.

Cholangiocytes express water channels (i.e. aquaporins (AQPs)), proteins that are increasingly recognized as important in water transport by biliary epithelia. However, direct functional studies demonstrating AQP-mediated water transport in cholangiocytes are limited, in part because of the lack of specific AQP inhibitors. To address this issue, we designed, synthesized, and utilized small interfering RNAs (siRNAs) selective for AQP1 and investigated their effectiveness in altering AQP1-mediated water transport in intrahepatic bile duct units (IBDUs) isolated from rat liver. Twenty-four hours after transfection of IBDUs with siRNAs targeting two different regions of the AQP1 transcript, both AQP1 mRNA and protein expression were inhibited by 76.6-92.0 and 57.9-79.4%, respectively. siRNAs containing the same percent of base pairs as the AQP1-siRNAs but in random sequence (i.e. scrambled siRNAs) had no effect. Suppression of AQP1 expression in cholangiocytes resulted in a decrease in water transport by IBDUs in response to both an inward osmotic gradient (200 mosm) or a secretory agonist (forskolin), the osmotic water permeability coefficient ( $P_f$ ) decreasing up to 58.8% and net water secretion ( $J_v$ ) decreasing up to 87%. A strong correlation between AQP1 protein expression and water transport in IBDUs transfected with AQP1-siRNAs was consistent with the decrease in water transport by IBDUs resulting from AQP1 gene silencing by AQP1-siRNAs. This study is the first to demonstrate the feasibility of utilizing siRNAs to specifically reduce the expression of AQPs in epithelial cells and provides direct evidence of the contribution of AQP1 to water transport by biliary epithelia.

PMID: 12468529 [PubMed - indexed for MEDLINE]



**Type IV collagenase (M(r) 72,000) expression in human prostate: benign and malignant tissue.**

Stearns ME, Wang M.

Department of Pathology, Medical College of Pennsylvania, Philadelphia 19129.

The expression of type IV collagenase (M(r) 72,000) has been examined in tissues from patients with benign prostatic hyperplasia (6 patients) and varying Gleason grades of malignant prostate cancer (18 patients). Immunoperoxidase labeling indicated that expression of the type IV collagenase was weak or nonexistent in benign tissue but consistently strong in the glandular and ductal epithelial cells of prostate tumors diagnosed at Gleason grades 1-8. In moderate to advanced cancer (i.e., Gleason grades 2 to 8), invasive tumor foci in the stromal tissue produced relatively modest amounts of type IV collagenase. The normal stromal tissue (i.e., fibroblasts) uniformly failed to produce detectable levels of type IV collagenase in the 24 patients examined. Northern and quantitative slot blot hybridization assays demonstrated that collagenase type IV mRNA levels were low in benign tissue and high in malignant tumors. In contrast, the stromal cells did not express significant amounts of type IV collagenase mRNA. Enzyme-linked immunosorbent assays demonstrated that the amounts of type IV collagenase protein correlated directly with the mRNA levels in the tumor tissue. The studies suggest that type IV collagenase may be selectively overexpressed by malignant, preinvasive prostatic epithelial cells.

PMID: 7679051 [PubMed - indexed for MEDLINE]



**The decompensated detrusor III: impact of bladder outlet obstruction on sarcoplasmic endoplasmic reticulum protein and gene expression.**

**Stein R, Gong C, Hutcheson JC, Canning DA, Zderic SA.**

Division of Urology, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA.

**PURPOSE:** Regulation of calcium ion homeostasis has a significant role in smooth muscle contractility. The sarcoplasmic endoplasmic reticulum, calcium, magnesium, adenosine triphosphatase (SERCA) is a regulatory ion pump that may have a role in the functional outcome after outlet obstruction. We investigate what correlation if any existed between SERCA protein and gene expression, and the contractile properties in the same bladder. **MATERIALS AND METHODS:** Standardized partial bladder outlet obstructions were created in adult New Zealand white rabbits, which were divided into control, sham operated and obstructed groups. Muscle strip studies subcategorized the obstructed group into compensated (force greater than 50% of control) and decompensated (force less than 50% of control). Microsomal membrane and total RNA fractions were prepared from the same bladder tissue. Membrane proteins were used for Western blot analysis using a SERCA specific monoclonal antibody, and total RNA was assessed with Northern blot analysis. **RESULTS:** The relative intensities of signals for the Western and Northern blots demonstrated a strong correlation between protein and gene expression. Furthermore there was a strong association between the loss of SERCA messenger RNA and protein expression and loss of bladder function. **CONCLUSIONS:** Bladder contractility after outlet obstruction is influenced in part by smooth muscle cell ability to maintain calcium homeostasis via SERCA. The loss of SERCA protein expression is mediated by down-regulation in gene expression in the same bladder. These data suggest that smooth muscle ion pump gene expression is in part mechanically (pressure work) regulated.

PMID: 10958733 [PubMed - indexed for MEDLINE]

**TNF-alpha and IL-8 are upregulated in the epidermis of normal human skin after UVB exposure: correlation with neutrophil accumulation and E-selectin expression.**

Strickland I, Rhodes LE, Flanagan BF, Friedmann PS.

Department of Dermatology, University of Liverpool, United Kingdom.

The in vivo response to ultraviolet B (UVB) radiation in skin is characterized by the accumulation of both mononuclear and polymorphonuclear cells within the dermis and an induction of vascular endothelial adhesion molecules. Epidermal production of cytokines (IL-8 and TNF-alpha) has been strongly implicated in the development of UVB-induced inflammation. In the current study, we examined the time course of IL-8 and TNF-alpha mRNA and protein expression in the epidermis over a 24-h period after in vivo UVB irradiation. Also, the induction of adhesion molecule expression and the accumulation of neutrophils within the dermis were followed. We found constitutive expression of both cytokines (mRNA and protein) in the epidermis of unirradiated skin. IL-8 was rapidly upregulated after irradiation and mRNA and protein increased at 4 h, reaching a maximum between 8 and 24 h. TNF-alpha mRNA and protein was minimally increased by 8 h after UVB irradiation and reached a maximum by 24 h. No significant alteration in ICAM-1 or VCAM-1 expression was observed. E-selectin expression, which was absent from control samples, was increased from 4 h onward and also reached a maximum at 24 h, coinciding with peak neutrophil accumulation. A strong correlation ( $r = 0.96$ ) was found between number of E-selectin-positive vessels and numbers of infiltrating neutrophils at this time. Moreover, because E-selectin expression was increased before any apparent increase in TNF-alpha protein (4 h), TNF-alpha does not appear to be involved in the early induction of the adhesion molecule, but cytokines such as TNF-alpha and IL-8 may act subsequently to augment the inflammatory response.

PMID: 9129230 [PubMed - indexed for MEDLINE]



**Basic fibroblast growth factor expression is increased in human renal fibrogenesis and may mediate autocrine fibroblast proliferation.**

Strutz F, Zeisberg M, Hemmerlein B, Sattler B, Hummel K, Becker V, Müller GA.

Department of Nephrology and Rheumatology, Georg-August-University Gottingen, Germany. fstrutz@gwdg.de

**BACKGROUND:** Interstitial fibroblasts play a critical role in renal fibrogenesis, and autocrine proliferation of these cells may account for continuous matrix synthesis. Basic fibroblast growth factor (FGF-2) is mitogenic for most cells and exerts intracrine, autocrine, and paracrine effects on epithelial and mesenchymal cells. The aims of the present studies were to localize and quantitate the expression of FGF-2 in normal and pathologic human kidneys and to study the *in vitro* effects of FGF-2 on proliferation, differentiation, and matrix production of isolated cortical kidney fibroblasts. **METHODS:** FGF-2 protein expression was localized by immunofluorescence double labelings in normal and fibrotic human kidneys. Subsequently, interstitial FGF-2 labeling was determined semiquantitatively in 8 normal kidneys and 39 kidneys with variable degrees of interstitial fibrosis and was correlated with the morphometrically determined interstitial cortical volume. In addition, FGF-2 expression was quantitated by immunoblot analysis in three normal and six fibrotic kidneys. FGF-2 mRNA was localized by *in situ* hybridizations. Seven primary cortical fibroblast lines were established, and expression of FGF-2 and FGF receptor-1 (FGFR-1) were examined. The effects of FGF-2 on cell proliferation were determined by bromodeoxyuridine incorporation and cell counts, those on differentiation into myofibroblasts by staining for alpha-smooth muscle actin, and those on matrix synthesis by enzyme-linked immunosorbent assay for collagen type I and fibronectin. Finally, proliferative activity *in vivo* was evaluated by expression of MIB-1 (Ki-67 antigen). **RESULTS:** In normal kidneys, FGF-2 expression was confined to glomerular, vascular, and a few tubular as well as interstitial fibroblast-like cells. The expression of FGF-2 protein was increased in human kidneys, with tubulointerstitial scarring correlating with the degree of interstitial fibrosis ( $r = 0.84$ ,  $P < 0.01$ ). Immunoblot analyses confirmed a significant increase in FGF-2 protein expression in kidneys with interstitial scarring. *In situ* hybridization studies demonstrated low-level detection of FGF-2 mRNA in normal kidneys. However, FGF-2 mRNA expression was robustly up-regulated in interstitial and tubular cells in end-stage kidneys, indicating that these cells are the source of excess FGF-2 protein. Primary cortical fibroblasts express FGF-2 and FGFR-1 *in vitro*. FGF-2 induced a robust growth response in these cells that could be blocked specifically by a neutralizing FGF-2 antibody. Interestingly, the addition of the neutralizing antibody alone did reduce basal proliferation up to 31.5%. In addition, FGF-2 induced expression of alpha-smooth muscle actin up to 1.6-fold, but no significant effect was observed on the synthesis of collagen type I and fibronectin. Finally, staining for MIB-1 revealed a good correlation of interstitial FGF-2 positivity

with interstitial and tubular proliferative activity ( $r = 0.71$ ,  $P < 0.01$  for interstitial proliferation,  $N = 30$ ). CONCLUSIONS: Interstitial FGF-2 protein and mRNA expression correlate with interstitial scarring. FGF-2 is a strong mitogen for cortical kidney fibroblasts and may promote autocrine fibroblast growth. Expression of FGF-2 correlates with interstitial and tubular proliferation in vivo.

PMID: 10760088 [PubMed - indexed for MEDLINE]



**Adiposity elevates plasma MCP-1 levels leading to the increased CD11b-positive monocytes in mice.**

Takahashi K, Mizuarai S, Araki H, Mashiko S, Ishihara A, Kanatani A, Itadani H, Kotani H.

Banyu Tsukuba Research Institute in collaboration with Merck Research Laboratories, Tsukuba, Ibaraki 300-2611, Japan.

Obesity is currently considered as an epidemic in the western world, and it represents a major risk factor for life-threatening diseases such as heart attack, stroke, diabetes, and cancer. Taking advantage of DNA microarray technology, we tried to identify the molecules explaining the relationship between obesity and vascular disorders, comparing mRNA expression of about 12,000 genes in white adipose tissue between normal, high fat diet-induced obesity (DIO) and d-Trp34 neuropeptide Y-induced obesity in mice. Expression of monocyte chemoattractant protein-1 (MCP-1) mRNA displayed a 7.2-fold increase in obese mice as compared with normal mice, leading to substantially elevated MCP-1 protein levels in adipocytes. MCP-1 levels in plasma were also increased in DIO mice, and a strong correlation between plasma MCP-1 levels and body weight was identified. We also showed that elevated MCP-1 protein levels in plasma increased the CD11b-positive monocyte/macrophage population in DIO mice. Furthermore, infusion of MCP-1 into lean mice increased the CD11b-positive monocyte population without inducing changes in body weight. Given the importance of MCP-1 in activation of monocytes and subsequent atherosclerotic development, these results suggest a novel role of adiposity in the development of vascular disorders.

PMID: 13129912 [PubMed - indexed for MEDLINE]



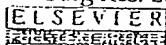
**Augmented expression of neuronal nitric oxide synthase in the atria parasympathetically decreases heart rate during acute myocardial infarction in rats.**

**Takimoto Y, Aoyama T, Tanaka K, Keyamura R, Yui Y, Sasayama S.**

Department of Cardiovascular Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan.

**BACKGROUND:** Nitric oxide (NO) synthesized within sinoatrial cells recently has been shown to participate in the autonomic control of heart rate. We hypothesized that NO in the neuronal cells in the heart was increased and parasympathetically regulated heart rate after myocardial infarction (MI). **METHODS AND RESULTS:** We examined heart rate dynamics and neuronal NO synthase (nNOS) expression and activities in the atria of rats with MI 1, 3, 7, and 14 days after MI (n=7 to 22 for each group). Both the mRNA levels of nNOS in the atria determined by competitive reverse transcriptase-polymerase chain reaction and the protein levels determined by Western blotting were significantly increased compared with controls 1, 3, and 7 days after MI. nNOS activity in the atria 1 day after infarction was also increased in MI rats. nNOS immunoreactivity was observed in nerve fibers in the atria. After infusion of a specific inhibitor of nNOS and iNOS, 1-(2-trifluoromethylphenyl) imidazole (TRIM) (50 mg/kg IV), heart rate was significantly ( $P<0.01$ ) increased in MI rats compared with controls 1, 3, and 7 days after MI. The iNOS-specific inhibitor, 1400W (10 mg/kg SC), did not significantly affect the heart rate in rats with MI. The effect of TRIM was abolished by pretreatment with L-arginine (25 mg/kg IV) or by parasympathetic blockade with atropine but not by propranolol. There was a strong correlation ( $r=0.837$ ,  $P<0.0001$ ) between the nNOS protein expression and heart rate change after TRIM infusion. **CONCLUSIONS:** These results indicate that increased nNOS parasympathetically decreased heart rate via the production of NO in rats with acute MI.

PMID: 11815433 [PubMed - indexed for MEDLINE]



# **Differential upregulation of cellular adhesion molecules at the sites of oxidative stress in experimental acute pancreatitis.**

**Telek G, Ducroc R, Scoazec JY, Pasquier C, Feldmann G, Roze C.**

INSERM U 410, Universite Paris 7 Denis Diderot, 75870 Paris, France.

**BACKGROUND:** Severe acute pancreatitis (AP)(2) is associated with exaggerated leukocyte adherence and activation. Endothelial cellular adhesion molecules (CAMs) can be induced by cytokines, but also directly by oxygen free radicals (OFRs), mediated by nuclear factor kappa-B (NF-kappa B). We investigated the behavior of inducible CAMs in relation to pancreatic oxidative stress. Our novel modification of cerium capture histochemistry (reaction of OFRs with cerium produces laser reflective Ce perhydroxide precipitates) combined with reflectance confocal laser scanning microscopy (CLSM) allows the histological codemonstration of in vivo OFR production and immunolabeled CAMs, or NF-kappa B. **METHODS:** Taurocholate AP was induced in rats; sham operated and normal animals served as controls. To achieve in situ, in vivo reaction of cerium with OFRs, animals were perfused with CeCl(3) solution at different time points (1, 2, 8, 24 h) and then sacrificed. E-selectin, P-selectin, ICAM-1, VCAM, and NF-kappa B p65 were labeled by immunofluorescence (IF) on frozen sections of cerium perfused pancreata. IF and Ce perhydroxide reflectance were simultaneously detected by CLSM. Pancreatic gene expression of the same CAMs was quantified by competitive RT-PCR (MIMIC internal control). **RESULTS:** Control pancreata showed negligible reflectance and minimal CAM expression. Early (1, 2 h) AP samples were characterized by intense, heterogeneous acinar OFR production, strong P-selectin, and increasing ICAM expression, with nuclear translocation of p65, histologically all colocalizing with the areas of acinar oxidative stress. Adherent polymorphonuclear leukocytes (PMNs) displayed weak OFR formation. Later (8, 24 h), a slowly declining P-selectin, but persisting ICAM-1 expression, was paralleled by widespread adherence of PMNs producing surprisingly large amounts of OFRs. VCAM and E-selectin showed a mild increase at 24 h. CAM gene activation was in good correlation with the protein expression. **CONCLUSIONS:** The early acinar oxidative stress is colocalized with NF-kappa B activation, preferential P-selectin, and ICAM upregulation in this AP model. Subsequently, adherent, activated PMNs become the major source of OFRs, thereby contributing to tissue damage. Copyright 2001 Academic Press.

PMID: 11180997 [PubMed - indexed for MEDLINE]



### **Myotonic dystrophy: an unstable CTG repeat in a protein kinase gene.**

**Timchenko L, Monckton DG, Caskey CT.**

Department of Molecular and Human Genetics, Baylor College of Medicine, Texas Medical Center, Houston 77030, USA.

Myotonic dystrophy (DM) is caused by the amplification of CTG repeats in the 3' untranslated region of a gene encoding a protein homologous to serine/threonine protein kinases. In DM patients the CTG repeats are extremely unstable, varying in length from patient to patient and generally increasing in length in successive generations. There is a strong correlation between the size of the repeats and the age of onset and severity of the disease. The molecular basis of the effect of the CTG expansion on the development of the DM phenotype continues to be investigated. The first working hypothesis of the molecular mechanism of DM was a reduction in steady-state myotonin-protein kinase (Mt-PK) mRNA and protein levels. However, although the consensus finding is that the Mt PK mRNA and protein levels are decreased in DM patients, it is still not clear if this reduction leads directly to the DM phenotype. In this short review we discuss the molecular aspects of CTG instability and the expression of the myotonin-protein kinase gene in normal and DM populations.

#### **Publication Types:**

- [Review](#)
- [Review, Tutorial](#)

PMID: 7620117 [PubMed - indexed for MEDLINE]

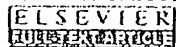
**Induction of class 3 aldehyde dehydrogenase in the mouse hepatoma cell line Hepa-1 by various chemicals.**

Torronen R, Korkalainen M, Karenlampi SO.

Department of Physiology, University of Kuopio, Finland.

The mouse hepatoma cell line Hepa-1 was shown to express an aldehyde dehydrogenase (ALDH) isozyme which was inducible by TCDD and carcinogenic polycyclic aromatic hydrocarbons. The induced activity could be detected with benzaldehyde as substrate and NADP as cofactor (B/NADP ALDH). As compared with rat liver and hepatoma cell lines, the response was moderate (maximally 5-fold). There was an apparent correlation between this specific form of ALDH and aryl hydrocarbon hydroxylase (AHH) in the Hepa-1 wild-type cell line--in terms of inducibility by several chemicals. However, the magnitude of the response was clearly smaller for ALDH than for AHH. Southern blot analysis showed that a homologous gene (class 3 ALDH) was present in the rat and mouse genome. The gene was also expressed in Hepa-1 and there was a good correlation between the increase of class 3 ALDH-specific mRNA and B/NADP ALDH enzyme activity after exposure of the Hepa-1 cells to TCDD. It is concluded that class 3 ALDH is inducible by certain chemicals in the mouse hepatoma cell line, although the respective enzyme is not inducible in mouse liver *in vivo*.

PMID: 1505055 [PubMed - indexed for MEDLINE]



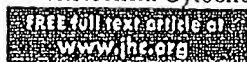
**Relationship between cyclin D1 and p21(Waf1/Cip1) during differentiation of human myeloid leukemia cell lines.**

**Ullmannova V, Stockbauer P, Hradcova M, Soucek J, Haskovec C.**

Department of Molecular Genetics, Institute of Hematology and Blood Transfusion, U Nemocnice 1, 128 20 Prague 2, Czech Republic. ullman@uhkt.cz

Expression of cell cycle-regulating genes was studied in human myeloid leukemia cell lines ML-1, ML-2 and ML-3 during induction of differentiation in vitro. Myelomonocytic differentiation was induced by phorbol ester (12-o-Tetradecanoyl-phorbol-13-acetate, TPA), tumor necrosis factor alpha (TNFalpha) or interferon gamma (INFgamma), or their combination. Differentiation (with the exception of TNFalpha alone) was accompanied by inhibition of DNA synthesis and cell cycle arrest. Inhibition of proliferation was associated with a decrease in the expression of cdc25A and cdc25B, cdk6 and Ki-67 genes, and with increased p21(Waf1/Cip1) gene expression, as measured by comparative RT-PCR. Expression of the following genes was not changed after induction of differentiation: cyclin A1, cyclin D3, cyclin E1 and p27(Kip1). Surprisingly, cyclin D1 expression was upregulated after induction by TPA, TNFalpha with INFgamma or BA. Cyclin D2 was upregulated only after induction by BA. The results of the expression of the tested genes obtained by comparative RT-PCR were confirmed by quantitative real-time (RQ) RT-PCR and Western blotting. Quantitative RT-PCR showed as much as a 288-fold increase of cyclin D1 specific mRNA after a 24h induction by TPA. The upregulation of cyclin D1 in differentiating cells seems to be compensated by the upregulation of p21(Waf1/Cip1). These results, besides others, point to a strong correlation between the expression of cyclin D1 and p21(Waf1/Cip1) on the one hand and differentiation on the other hand in human myeloid leukemic cells and reflect a rather complicated network regulating proliferation and differentiation of leukemic cells.

PMID: 12921950 [PubMed - indexed for MEDLINE]



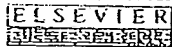
**Intestinal carbamoyl phosphate synthase I in human and rat. Expression during development shows species differences and mosaic expression in duodenum of both species.**

Van Beers EH, Rings EH, Posthuma G, Dingemanse MA, Taminiou JA, Heymans HS, Einerhand AW, Buller HA, Dekker J.

Pediatric Gastroenterology and Nutrition, Department Pediatrics, Emma Children's Hospital, Academic Medical Center, Amsterdam, The Netherlands.

The clinical importance of carbamoyl phosphate synthase I (CPSI) relates to its capacity to metabolize ammonia, because CPSI deficiencies cause lethal serum ammonia levels. Although some metabolic parameters concerning liver and intestinal CPSI have been reported, the extent to which enterocytes contribute to ammonia conversion remains unclear without a detailed description of its developmental and spatial expression patterns. Therefore, we determined the patterns of enterocytic CPSI mRNA and protein expression in human and rat intestine during embryonic and postnatal development, using in situ hybridization and immunohistochemistry. CPSI protein appeared during human embryogenesis in liver at 31-35 e. d. (embryonic days) before intestine (59 e.d.), whereas in rat CPSI detection in intestine (at 16 e.d.) preceded liver (20 e.d.). During all stages of development there was a good correlation between the expression of CPSI protein and mRNA in the intestinal epithelium. Strikingly, duodenal enterocytes in both species exhibited mosaic CPSI protein expression despite uniform CPSI mRNA expression in the epithelium and the presence of functional mitochondria in all epithelial cells. Unlike rat, CPSI in human embryos was expressed in liver before intestine. Although CPSI was primarily regulated at the transcriptional level, CPSI protein appeared mosaic in the duodenum of both species, possibly due to post-transcriptional regulation.

PMID: 9446830 [PubMed - indexed for MEDLINE]



**Expression of deoxycytidine kinase in leukaemic cells compared with solid tumour cell lines, liver metastases and normal liver.**

van der Wilt CL, Kroep JR, Loves WJ, Rots MG, Van Groenigen CJ, Kaspers GJ, Peters GJ.

Department of Medical Oncology, VU University Medical Center, Amsterdam, The Netherlands.

Deoxycytidine kinase (dCK) is required for the phosphorylation of several deoxyribonucleoside analogues that are widely employed as chemotherapeutic agents. Examples include cytosine arabinoside (Ara-C) and 2-chlorodeoxyadenosine (CdA) in the treatment of acute myeloid leukaemia (AML) and gemcitabine to treat solid tumours. In this study, expression of dCK mRNA was measured by a competitive template reverse transcriptase polymerase chain reaction (CT RT-PCR) in seven cell lines of different histological origin, 16 childhood and adult AML samples, 10 human liver samples and 11 human liver metastases of colorectal cancer origin. The enzyme activity and protein expression levels of dCK in the cell lines were closely related to the mRNA expression levels ( $r=0.75$ ,  $P=0.026$  and  $r=0.86$ ,  $P=0.007$ ). In AML samples, dCK mRNA expression ranged from 1.16 to 35.25 ( $\times 10^{-3}$ ) dCK/beta-actin. In the cell line panel, the range was 2.97-56.9 ( $\times 10^{-3}$ ) dCK/beta-actin of dCK mRNA expression. The enzyme activity in liver metastases was correlated to dCK mRNA expression ( $r=0.497$ ,  $P=0.05$ ). In the liver samples, these were not correlated. dCK mRNA expression showed only a 36-fold range in liver while a 150-fold range was observed in the liver metastases. In addition, dCK activity and mean mRNA levels were 2.5-fold higher in the metastases than in the liver samples. Since dCK is associated with the sensitivity to deoxynucleoside analogues and because of the good correlation between the different dCK measurements in malignant cells and tumours, the CT-RT PCR assay will be useful in the selection of patients that can be treated with deoxycytidine analogues.

PMID: 12628850 [PubMed - indexed for MEDLINE]

### **Expression of cytokines and growth factors in human glomerulonephritides.**

**Waldherr R, Noronha IL, Niemir Z, Kruger C, Stein H, Stumm G.**

Department of Pathology, University of Heidelberg, Germany.

Numerous experimental studies point to the potential role of cytokines and growth factors in the pathogenesis of renal disease. However, from the various autocrine and paracrine mediators identified in vitro and in animal models, so far only a few have been demonstrated in selected human glomerulopathies. We examined two types of glomerulonephritis (GN): extracapillary GN with anti-neutrophil cytoplasmic autoantibodies (ANCA), an example of an acute form of GN, and mesangial IgA GN, usually a chronic form of GN, with immunocytochemistry, in situ hybridization and the polymerase chain reaction. Normal renal tissue from tumour nephrectomies served as a control. In ANCA-positive GN with active renal lesions (crescents, glomerular and vascular necrosis), infiltrating mononuclear cells in glomeruli and in the interstitium expressed interleukin (IL)-1 beta, tumour necrosis factor (TNF)-alpha, IL-2, interferon (IFN)-gamma, platelet-derived growth factor (PDGF) and transforming growth factor (TGF)-beta. Cytokine expression was also observed in activated resident cells, including endothelial cells, capsular epithelial cells, smooth muscle cells of vessel walls, fibroblasts and some tubular epithelial cells. In addition, we noted an increase in the cytokine and growth factor receptors TNF-R, IL-1R type II, IL-2R, IFN-gamma R and PDGF beta-R. In contrast, in mesangial IgA-GN, IL-1 beta, TNF-alpha, IFN-gamma and IL-2 were usually absent in glomeruli. Mesangial expansion in this disorder was accompanied by an increased expression of PDGF, PDGF beta-R, TGF-beta and IL-6 in mesangial areas. In both conditions a good correlation was observed between cytokine expression at the mRNA (in situ hybridization) and protein level (immunocytochemistry). (ABSTRACT TRUNCATED AT 250 WORDS)

#### **Publication Types:**

- [Review](#)
- [Review, Tutorial](#)

PMID: 8398664 [PubMed - indexed for MEDLINE]

**Malignant transformation of the human endometrium is associated with overexpression of lactoferrin messenger RNA and protein.**

Walmer DK, Padin CJ, Wrona MA, Healy BE, Bentley RC, Tsao MS, Kohler MF, McLachlan JA, Gray KD.

Department of Obstetrics and Gynecology, Duke University Medical Center, Durham, North Carolina 27710.

In the mouse uterus, lactoferrin is a major estrogen-inducible uterine secretory protein, and its expression correlates directly with the period of peak epithelial cell proliferation. In this study, we examine the expression of lactoferrin mRNA and protein in human endometrium, endometrial hyperplasias, and adenocarcinomas using immunohistochemistry, Western immunoblotting, and Northern and in situ RNA hybridization techniques. Our results reveal that lactoferrin is expressed in normal cycling endometrium by a restricted number of glandular epithelial cells located deep in the zona basalis. Two thirds (8 of 12) of the endometrial adenocarcinomas examined overexpress lactoferrin. This tumor-associated increase in lactoferrin expression includes an elevation in the mRNA and protein of individual cells and an increase in the number of cells expressing the protein. In comparison, only 1 of the 10 endometrial hyperplasia specimens examined demonstrates an increase in lactoferrin. We also observe distinct cytoplasmic and nuclear immunostaining patterns under different fixation conditions in both normal and malignant epithelial cells, similar to those previously reported in the mouse reproductive tract. Serial sections of malignant specimens show a good correlation between the localization of lactoferrin mRNA and protein in individual epithelial cells by in situ RNA hybridization and immunohistochemistry. Although the degree of lactoferrin expression in the adenocarcinomas did not correlate with the tumor stage, grade, or depth of invasion in these 12 patients, there was a striking inverse correlation between the presence of progesterone receptors and lactoferrin in all 8 lactoferrin-positive adenocarcinomas. In summary, lactoferrin is expressed in a region of normal endometrium known as the zona basalis which is not shed with menstruation and is frequently overexpressed by progesterone receptor-negative cells in endometrial adenocarcinomas.

PMID: 7867003 [PubMed - indexed for MEDLINE]

TPL



# Cancer Research

AN OFFICIAL JOURNAL OF THE AMERICAN ASSOCIATION FOR CANCER RESEARCH



March 1, 1995  
Volume 55 • Number 5  
PP. 975-1197  
ISSN 0008-5472 • CNREA 8

UNIVERSITY OF CALIFORNIA

MAR 6 1995

LIBRARY OF THE





# Cancer Research

AN OFFICIAL JOURNAL OF THE AMERICAN ASSOCIATION FOR CANCER RESEARCH

**American Association for Cancer Research, Inc.**  
Publications Staff  
Managing Editor  
Margaret Foti  
Assistant Managing Editor  
Mary Anne Menon  
Manager, Editorial Services  
Heide M. Pustay  
Staff Editors  
Lisa A. Chippendale  
Patricia M. Grchow  
Michael J. Beveridge  
Kathleen C. Assenmacher  
Supervisor, Manuscript Processing  
Margaret A. Pickels  
Assistant Supervisor, Manuscript Processing  
Theresa A. Griffith  
Staff Assistant  
Mary Ellen Furing  
Editorial Assistants  
Valerie L. Taylor  
Andrea Conrad  
Katherine V. Pawlowaki  
Bethann Massarella  
Charles W. Wells, Jr.  
Administrative Staff  
Executive Director  
Margaret Foti  
Director of Administration  
Adam D. Blisstein  
Controller  
Joan D. Brokenshire  
Coordinator, Financial Operations  
George L. Moore  
Meetings Coordinator  
Jeffrey M. Ruben  
Meeting Planner  
Carol L. Kanoff  
Public Information Coordinator  
Jenny Anne Hori-Mertz  
Executive Assistant for Programs  
Ruth E. Fortson  
Administrative Assistant  
Robin E. Felder  
Systems Specialist  
Lydia I. Rodriguez  
Staff Assistant  
Lori L. Holmes  
Secretary  
Malika L. Wright  
Editorial Secretary  
Diana F. Certo  
Financial Clerk  
Paul Penzault  
Office Clerk  
James J. Waters  
Data Entry Clerk  
Robert A. Simms II

Editor-in-Chief  
Carlo M. Croce

**Associate Editors**

Stuart A. Aaronson  
Jerry M. Adams  
David S. Alberts  
Carmen J. Allegra  
Frederick R. Appelbaum  
James O. Armstrong  
William M. Baird  
Allan Balmain  
J. Carl Barrett  
Renato Baserga  
William T. Beck  
Joseph R. Bertino  
Mina J. Bissell  
Clara D. Bloomfield  
Gianni Bonadonna  
Ernest C. Borden  
G. Tim Bowden  
Edward Bresnick  
Samuel Broder  
Ronald N. Buick  
Paul A. Bunn, Jr.  
C. Patrick Burns  
Fernando Cabanillas  
Bruno Calabretta  
Eli Casanovi  
Robert L. Capizzi  
Weizhi K. Cavenee  
Bruce D. Cheson  
David Colcher  
Robert L. Comis  
Allan H. Conney  
Neal G. Copeland  
Joseph G. Cory  
William M. Crist  
Benoit de Crombrughe

Riccardo Dalla-Favera  
Richard L. Davidson  
Eugene R. DeSombre  
T. Michael Dexter  
Ralph E. Durand  
Alan Eastman  
Gertrude B. Elion  
Leonard C. Erickson  
Nelson Fausto  
Eric R. Fearon  
Isaiah J. Fidler  
Robert A. Floyd  
Jude Folkman  
Kenneth A. Foon  
Joseph F. Fraumeni, Jr.  
Stephen H. Friend  
Minoru Fukuda  
Philip Furmanski  
Eugene W. Gerner  
Eli Gluskin  
David W. Golde  
Michael M. Gottesman  
J. W. Grisham  
F. Peter Guentrich  
Sen-irob Hakomori  
Phillip C. Hanawalt  
Curtis C. Harris  
John C. Harshbarger  
Stephen S. Hecht  
Carl-Henrik Heldin  
Brian E. Henderson  
Richard B. Hochberg  
Wau Ki Hong  
Tasuku Honjo  
Kay Huebner

John T. Isaacs  
Mark A. Israel  
Elaine S. Jaffe  
Rakesh K. Jain  
Peter A. Jones  
V. Craig Jordan  
John H. Kersey  
Young S. Kim  
Kenneth W. Kinzler  
Tadamitsu Kishimoto  
Alfred G. Knudson, Jr.  
Kurt W. Kohle  
Lawrence N. Kolonel  
Donald W. Kufe  
John S. Lazo  
Jay A. Levy  
Frederick P. Li  
Victor Ling  
Lance A. Lione  
Martin Lipkin  
Marc E. Lippman  
Gerald Lincoff  
Leroy F. Liu  
Lawrence A. Loeb  
Dan L. Longo  
Reuben Lotan  
David B. Ludlum  
Tak W. Mak  
Michael J. Mastrangelo  
Lynn M. Maurisano  
W. Gillies McKenna  
Anna T. Meadows  
John Mendelsohn  
Christopher J. Michiels  
John D. Minna

Beatrice Mintz  
Malcolm S. Mitchell  
Ruth J. Muschel  
Yusuke Nakamura  
Garth L. Nicolson  
Kenneth Nilsson  
Susumu Nishimura  
Jeffrey A. Norton  
Kenneth Olden  
Gilbert S. Omenn  
Richard J. O'Reilly  
Robert F. Ozols  
Michael A. Palladin, Jr.  
Peter L. Pedersen  
Anthony E. Pegg  
Angel Pellicer  
Bice Perussia  
Gordon L. Phillips  
Cecil B. Picken  
Jacalyn H. Pierce  
Vito Quaranta  
Frank J. Rauscher III  
Donald J. Reed  
Ralph A. Reisfeld  
Richard A. Rifkin  
Leslie L. Robinson  
Igor B. Roninson  
Howard I. Scher  
Robert T. Schimke  
J. Schlessinger  
Manfred Schwab  
Joseph V. Simone  
Francis M. Sirotnak  
Anna Marie Skala

Michael B. Sporn  
Martha R. Stampfer  
Eric J. Stanbridge  
Gary S. Stein  
Bernard S. Strauss  
Takashi Sugimura  
Saraswati Sukumar  
James A. Swenberg  
Paul Talalay  
Tadatsugu Taniguchi  
Steven R. Tannenbaum  
Masaki Terada  
Kenneth D. Tew  
Donald J. Tindall  
George J. Todaro  
Jeffrey M. Trent  
Giorgio Trinchieri  
Takashi Tsuruo  
Axel Ullrich  
Peter Vaupel  
Giancarlo Vecchio  
Daniel D. Von Hoff  
Michael D. Waterfield  
Lee W. Watanberg  
Ralph R. Weichselbaum  
I. Bernard Weinstein  
Bengt Westermark  
Raymond L. White  
Gordon F. Whitmore  
Max S. Wicha  
Walter Willett  
H. Rodney Withers  
Sheldon Wolff  
Stuart H. Yuspa  
Harald zur Hausen

**Cover Editorial Board**  
Sidney Weinhouse,  
Cover Editor

Hugh J. Ceech  
Clark W. Heath, Jr.

Edwin A. Mirand  
Raymond W. Ruddon

Takashi Sugimura  
John H. Weisburger

*Cancer Research* is sponsored by the American Association for Cancer Research, Inc. and receives major support from the American Cancer Society, Inc. Publication costs are also met by a grant from the Elsa U. Pardee Foundation. Accelerated mailing of journals to AACR members in Japan is supported by a generous grant from the Banyu Pharmaceutical Company.

**Subscription Information**

*Cancer Research* is published twice a month, one volume per year, by the American Association for Cancer Research, Inc. (AACR). Subscriptions include the *Proceedings of the American Association for Cancer Research*, issued in March of each year. Except for members of the AACR, all subscriptions are payable in advance to *Cancer Research*, P.O. Box 5000, Denville, NJ 07834 [Telephone: (800) 875-2997; FAX: (201) 627-5872], to which all business communications, remittances (in United States currency or its equivalent), and subscription orders should be sent. In Japan, send orders and inquiries to (sole agent): USACO Corporation, Toyosumi Bldg., 15-12, Shimobashi 1-chome, Minato-ku, Tokyo 105, Japan; Tel. (03) 502-6471. Individuals who are not AACR members may subscribe to Volume 55 (1995) of *Cancer Research* at the rate of \$460 U.S./\$520 foreign. *Cancer Research* is only available to institutions as a combined subscription with *Clinical Cancer Research*. The combined 1995 institutional subscription price of \$495 U.S./\$575 foreign includes a subscription to *Clinical Cancer Research*. Canadian subscribers should add 7% GST. Changes of address notification should be sent 60 days in advance and include both old and new addresses. Member subscribers should send address changes to: AACR Member Services, Public Ledger Bldg., Suite E16, 150 South Independence Mall West, Philadelphia, PA 19106-3483. Nonmember subscribers should send changes of address to: *Cancer Research*, P.O. Box 5000, Denville, NJ 07834. Copies of the journal which are undeliverable because of address changes will be destroyed.

# Malignant Transformation of the Human Endometrium Is Associated with Overexpression of Lactoferrin Messenger RNA and Protein

David K. Walmer,<sup>1</sup> Cheryl J. Padin, Mark A. Wrona, Bridget E. Healy, Rex C. Bentley, Ming-Sound Tsao, Matthew F. Kohler, John A. McLachlan, and Karen D. Gray

Department of Obstetrics and Gynecology, Duke University Medical Center, Durham, North Carolina 27710 [D. K. W., B. E. H., M. F. K.]; Department of Obstetrics and Gynecology, University of Michigan, Ann Arbor, Michigan 48109 [C. J. P.]; School of Medicine, University of Maryland, Baltimore, Maryland 21201 [M. A. W.]; Department of Pathology, Duke University Medical Center, Durham, North Carolina 27710 [R. C. B.]; Department of Pathology, Montreal General Hospital, Montreal, Quebec, Canada H3G1A4 [M.-S. T.]; Laboratory of Reproductive and Developmental Toxicology, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709 [J. A. M.]; and Department of Obstetrics and Gynecology, Uniformed Services University of the Health Sciences, Bethesda, Maryland 20814-7799 [K. D. G.]

## ABSTRACT

In the mouse uterus, lactoferrin is a major estrogen-inducible uterine secretory protein, and its expression correlates directly with the period of peak epithelial cell proliferation. In this study, we examine the expression of lactoferrin mRNA and protein in human endometrium, endometrial hyperplasias, and adenocarcinomas using immunohistochemistry, Western immunoblotting, and Northern and *in situ* RNA hybridization techniques. Our results reveal that lactoferrin is expressed in normal cycling endometrium by a restricted number of glandular epithelial cells located deep in the zona basalis. Two thirds (8 of 12) of the endometrial adenocarcinomas examined overexpress lactoferrin. This tumor-associated increase in lactoferrin expression includes an elevation in the mRNA and protein of individual cells and an increase in the number of cells expressing the protein. In comparison, only 1 of the 10 endometrial hyperplasia specimens examined demonstrates an increase in lactoferrin. We also observe distinct cytoplasmic and nuclear immunostaining patterns under different fixation conditions in both normal and malignant epithelial cells, similar to those previously reported in the mouse reproductive tract. Serial sections of malignant specimens show a good correlation between the localization of lactoferrin mRNA and protein in individual epithelial cells by *in situ* RNA hybridization and immunohistochemistry. Although the degree of lactoferrin expression in the adenocarcinomas did not correlate with the tumor stage, grade, or depth of invasion in these 12 patients, there was a striking inverse correlation between the presence of progesterone receptors and lactoferrin in all 8 lactoferrin-positive adenocarcinomas. In summary, lactoferrin is expressed in a region of normal endometrium known as the zona basalis which is not shed with menstruation and is frequently overexpressed by progesterone receptor-negative cells in endometrial adenocarcinomas.

## INTRODUCTION

The uterus is a sex steroid-responsive organ that plays a major role in women's health. Hysterectomies were the most frequently performed major surgical procedures in a 20-year study interval (1965-1984; Ref. 1). Fifty-eight to 80% of these 12.5 million procedures were performed for estrogen-related disorders of proliferation. Chronic unopposed estrogen exposure, most commonly associated with type II ovulatory disorders, eventually leads to the development of complex endometrial hyperplasia and adenocarcinoma. Since the sex steroids, estrogen and progesterone, act on their target tissues by regulating the expression of a wide variety of signaling molecules, identifying these regulatory factors will provide critical information towards understanding normal reproduction and reproductive tract pathology. Our current knowledge of estrogen and progesterone action on the reproductive tract is based to a great extent on information collected from rodents (2). Although differences exist

between the reproductive physiology of rodents and humans, the mouse has been a useful model for studying steroid hormone action in the human female reproductive tract (3, 4). One potential regulatory molecule shown to be regulated by estrogen in the mouse reproductive tract is lactoferrin. Lactoferrin is a basic glycoprotein with an extraordinarily high affinity for iron that was originally discovered in milk. This protein is expressed in a wide variety of tissues, most notably in polymorphonuclear leukocytes and most mammalian exocrine glandular secretions. In the mammary gland (5) and the female reproductive tract of the mouse (6-8), lactoferrin is regulated by endocrine hormones. Prolactin stimulates lactoferrin synthesis in the breast; whereas in the uterus and vagina, the ovarian sex steroid, 17 $\beta$ -estradiol, is the inducer. (6, 7, 9). To date, lactoferrin is one of the few genes that have been identified in the rodent that are directly regulated by estradiol. The lactoferrin gene contains an ERE<sup>2</sup> that is important for regulating its expression *in vivo* in the mouse reproductive tract. Being linked to estradiol, the expression of lactoferrin by the uterine epithelium parallels the onset of DNA synthesis. Although sequencing information suggests that the human lactoferrin gene also contains a functional imperfect ERE in the 5'-flanking promoter region (10, 11), there is very little data regarding lactoferrin expression in the human female reproductive tract.

The purpose of our study was to examine the expression of lactoferrin in the human endometrium under normal and pathological conditions by immunohistochemistry, immunoblotting, and Northern and *in situ* RNA hybridization techniques. In addition, we looked for correlations between lactoferrin expression and several parameters, such as the stage of the menstrual cycle, the distribution of estrogen and progesterone receptors, HER-2/*neu* expression, markers of cell proliferation, and the histopathological grade and extent of myometrial invasion in the adenocarcinomas. Our data demonstrates that lactoferrin is expressed in a very restricted number of glands in the basal region of normal human endometrium and is markedly overexpressed in a significant number of the uterine adenocarcinomas by PR-negative cells.

## MATERIALS AND METHODS

**Tissue Preparation and Histological Evaluation.** Surgical pathology specimens were obtained from Duke University Medical Center (Durham, NC) and the Department of Pathology at Montreal General Hospital (Quebec, Montreal, Canada). Cycling endometrium was obtained from 22 women (ages 31-49), and atrophic endometrium was obtained from 7 postmenopausal women (ages 64-77). Hysterectomies from cycling women were performed for subserosal leiomyomas ( $n = 6$ ), pelvic relaxation ( $n = 8$ ), pelvic pain ( $n = 4$ ), peritoneal endometriosis ( $n = 2$ ), and cancer of either the cervix ( $n = 1$ ) or the ovary ( $n = 1$ ). In addition, 12 adenocarcinomas, 3 atypical complex hyperplasias, 5 complex hyperplasias without atypia, and 4 simple hyperplasias were analyzed. Each human uterus was bivalved shortly after

Received 8/15/94; accepted 12/30/94.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

<sup>1</sup> To whom requests for reprints should be addressed, at Department of Obstetrics and Gynecology, Box 3143, Duke University Medical Center, Durham, NC 27710.

<sup>2</sup> The abbreviations used are: ERE, estrogen response element; PCNA, proliferating cell nuclear antigen; ER, estrogen receptor; PR, progesterone receptor.

hysterectomy, and endometrium was removed from the fundal region. A full thickness biopsy was placed into either 10% neutral-buffered formalin or Bouin's solution overnight at room temperature before dehydration, paraffin embedding, and sectioning at 4  $\mu$ m on silanized slides. Histological evaluations of hematoxylin and eosin-stained slides were performed blindly by one board-certified pathologist. Normal endometrial samples were dated by the criteria of Noyes *et al.* (12). Endometrial hyperplasias and carcinomas were classified according to the current recommendations, of the International Society of Gynecological Pathologists under the auspices of WHO (13). Histological grading of tumors was performed according to Federation Internationale des Gynecologues et Obstetristes criteria (14). Each specimen was read a minimum of three times, and only specimens that were read consistently the same way were included in the study. Unstained sections of the same tissues were used for the cytochemical analysis of protein and mRNA expression using specific reagents. A few endometrial samples were frozen for subsequent protein and RNA extraction, which were evaluated by Western and Northern blotting, respectively. All human tissues were handled with the precautions and the guidelines required by Duke University and National Institute of Environmental Health Sciences.

**Immunolocalization.** Slides chosen for study were deparaffinized and rinsed in 20% glacial acetic acid at 4°C for 15 s to inhibit endogenous alkaline phosphatase. All subsequent incubations and washes were at room temperature. Sections were next equilibrated in PBS for 20 min and incubated for 20 min with 1.5% normal goat serum diluted in PBS to block nonspecific binding. Detection of lactoferrin was performed primarily with a rabbit anti-human lactoferrin polyclonal antiserum generated in our laboratory and affinity purified. Similar results were also seen with a nonaffinity-purified commercial antiserum (Biogenex, San Ramon, CA). Following incubation at room temperature with primary antisera for 60 min, the sections were washed in PBS twice for 10 min each, and lactoferrin was localized using an alkaline phosphatase-biotin-superavidin detection system (Vectastain ABC-AP kit; Vector Laboratories, Burlingame, CA) or the Super Sensitive Detection System (Biogenex, San Ramon, CA). To identify nonspecific staining, preimmune rabbit serum was used in place of the primary antibody. The immunoreaction was quantitated by determining the percentage of glands and the percentage of cells staining for lactoferrin in the zona basalis and the zona functionalis, with a minimum of 300 cells counted in each region. PR antibody was provided by Geoffrey Greene (KID68), and a commercial source was also used (Biogenex, San Ramon, CA). Identical staining patterns were confirmed with both preparations. Other commercially obtained antisera include PCNA (Biogenex, San Ramon, CA), ER (ER1D5; AMAC, Westbrook, MA), MIB-1 (AMAC), and HER-2/*neu* (Biogenex, San Ramon, CA). The primary antisera incubations were 2 h for the PR, 1 h for PCNA, MIB-1, and ER, and 30 min for HER-2/*neu*. Antisera dilutions were 1:100 for MIB-1 and 1:20 for HER-2/*neu*. Antigen retrieval (Biogenex, San Ramon, CA) was performed before adding the progesterone primary antisera.

**Western Blot Analysis.** Proteins were extracted from endometrial biopsies by homogenization on ice in 1% Triton-X and 20 mM Tris-HCl (pH 7.4) with protease inhibitors (10  $\mu$ g/ml leupeptin, 200 KU/ml aprotinin, and 20  $\mu$ g/ml phenylmethylsulfonyl fluoride) and clarified by centrifugation at 45,000 rpm for 30 min in a Beckman 70.1 Ti rotor; then the supernatant was analyzed for protein concentration by the BCA protein assay (Pierce, Rockford, IL). Aliquots of 200  $\mu$ g were separated by electrophoresis on an 8.5% SDS polyacrylamide gel, blotted onto nitrocellulose membranes, incubated with polyclonal rabbit antihuman lactoferrin antisera, and localized with an  $^{125}$ I-labeled donkey anti-rabbit immunoglobulin, as described previously (7).

**In Situ Hybridization.** All slides were pretreated with 0.2 M HCl for 30 min at room temperature, digested with 1  $\mu$ g/ml proteinase-K (Sigma Chemical Co., St. Louis, MO) in 0.1 M Tris-HCl (pH 7.4)-0.05 M EDTA for 15 min at 37°C, and then treated with 0.1 M ribonuclease-0.25% acetic anhydride for 5 min at room temperature and 0.1 M Tris-glycine (pH 7.4) for 30 min at room temperature. The sections were subsequently dehydrated with graded ethanol, air dried, and prehybridized at 50°C for 1 h in 2  $\times$  SSC, 10 mM DTT, 5  $\times$  Denhardt's solution, 100  $\mu$ g/ml of both salmon sperm DNA and yeast tRNA, and 50% formamide (15). The slides were then hybridized overnight at 50°C in the same medium with 10% dextran sulfate and 2  $\times$  10<sup>6</sup> cpm/ $\mu$ l of the specific RNA probe. The lactoferrin oligonucleotide probe was amplified by PCR using primers that spanned nucleotides 718-1654 (10) and cloned into pGEM-42.  $^{32}$ S-labeled sense and antisense RNA probes were made with the

Promega Riboprobe kit (Promega, Madison, WI), washed twice in 1  $\times$  SSC for 10 min at room temperature, digested with RNase [2.8  $\mu$ g/ml RNase-A, 0.3  $\mu$ g/ml RNase-T1, 10 mM Tris-HCl (pH 7.4), and 15 mM NaCl], and washed again with 1  $\times$  SSC twice for 20 min each time at 50°C, twice for 20 in 0.1  $\times$  SSC at 55°C, and once for 20 min at 60°C. The sections were then dehydrated and dipped in Kodak autoradiographic emulsion (NTB-2) for detection of specific mRNA expression. The slides were allowed to develop for 2 weeks. After this period, the slides were developed using Kodak D19 developer and Kodak Rapid Fixer.

**Northern RNA Analysis.** Total cellular RNA was purified from frozen tissue by the guanidinium isothiocyanate-caesium chloride method, and poly(A<sup>+</sup>) RNA was isolated by oligo(dT)-cellulose chromatography using methods described previously (15). For Northern blot analysis, poly(A<sup>+</sup>) RNA was resolved by electrophoresis on 1.5% formaldehyde agarose gels, stained with ethidium bromide, and transferred to a nylon membrane. The membrane was probed with a  $^{32}$ P-labeled lactoferrin cDNA derived from human uterus (nucleotides 718-1654; accession no. S52659) using PCR techniques, followed by cloning into pGEM-42 (Promega, Madison, WI). In order to insure that the quality and quantity of RNA analyzed by Northern blotting was equivalent between control and treated groups, the blot was probed simultaneously for glyceraldehyde-3-phosphate dehydrogenase.

**Statistical Analysis.** Values are presented as means  $\pm$  SD. Differences between the zona basalis and functionalis were tested by the two-tailed Student's *t* test.

## RESULTS

### Immunohistochemical Analysis of Lactoferrin Protein

**Normal Cycling Endometrium.** Immunohistochemical studies of normal cycling human endometrium localize lactoferrin protein predominantly to the glandular epithelium deep in the zona basalis and not to the functionalis (Fig. 1A). The association of lactoferrin protein expression with the zona basalis is statistically significant ( $P < 0.001$ ; Fig. 2). Two to 56% of the glands express lactoferrin at any given time during the menstrual cycle. Within positive glands, lactoferrin protein immunolocalization is heterogeneous in that positively staining epithelial cells are interspersed with cells negative for lactoferrin expression. No apparent differences in morphology, PCNA, ER, or PR expression are seen to account for the heterogeneous pattern of intra- and intergland lactoferrin expression in normal endometrium. Similar to our previous findings in mouse uterine epithelial cells, the positive-staining glandular cells of the human endometrium demonstrate two distinct immunostaining patterns for lactoferrin, cytoplasmic and nuclear (Fig. 1B), seen with both formalin and Bouin's fixation. In evaluating the temporal expression of lactoferrin, there is a trend towards more glands expressing lactoferrin during the secretory phase. Because of the large variance, the trend is not statistically significant. As expected, the polymorphonuclear leukocytes in the endometrium also stain intensely for lactoferrin protein, which is stored in their secondary granules (Fig. 3). These results demonstrate, for the first time, that lactoferrin protein is expressed in the human endometrium predominantly by polymorphonuclear leukocytes and epithelial cells of glands located deep in the zona basalis.

**Proliferative Endometrial Disorders: Hyperplasias and Adenocarcinomas.** Immunohistochemistry reveals that the expression of lactoferrin protein is increased in 66.6% (8 of 12) of the endometrial adenocarcinomas examined. In one-half of these cases (4 of 12), lactoferrin is intensely expressed by malignant epithelial cells throughout the entire tumors (Fig. 4). The other four adenocarcinomas demonstrate increased staining for lactoferrin in concentrated regions of the tumors. In all eight cases where lactoferrin expression is elevated, the cells expressing lactoferrin have one similarity with normal positive glands in that they demonstrate heterogeneous staining of interspersed positive and negative cells. However, the expression of lactoferrin by the malignant cells clearly differs from normal

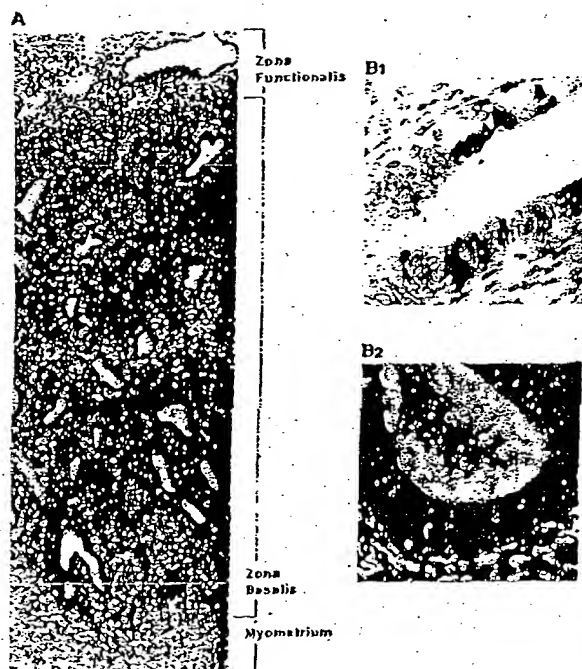


Fig. 1. Localization of lactoferrin protein in normal cycling endometrium by immunohistochemistry using a specific polyclonal antibody. Our analysis reveals that lactoferrin protein is present in a limited number of glands located in the zona basalis of the endometrium (A)  $\times 10$ . Also, note that lactoferrin is heterogeneous within positive glands, i.e., cells staining for lactoferrin are interspersed with negative-staining epithelial cells throughout the gland. Two immunohistochemical staining patterns are noted for lactoferrin in normal uterine epithelium. In one pattern, lactoferrin protein is immunolocalized primarily over the cytoplasm (B1), and in the other, the staining is seen over the nucleus (B2);  $\times 40$ .

positive glands in that the lactoferrin is not limited to the basal regions of the tumors, many more cells are positive, and the relative intensity of the staining over individual cells is increased. Although increased lactoferrin expression is associated with malignant transformation, we do not find a correlation between lactoferrin protein presence and the stage, nuclear grade, Fédération Internationale des Gynécologues et Obstétristes grade, or the depth of myometrial invasion in the 12 tumors studied (Table 1). In sharp contrast to the common dysregulation of lactoferrin expression found in the malignant endometrium, only 1 of 10 endometrial hyperplasia specimens evaluated contained an increased number of cells staining for lactoferrin. The hyperplastic specimen overexpressing lactoferrin was read as complex without atypia.

#### *In Situ* and Northern Analyses of Lactoferrin mRNA Expression in Normal and Malignant Endometrium

To further our understanding of the location of lactoferrin protein synthesis in the human endometrium, we examined lactoferrin mRNA expression by *in situ* and Northern hybridization using specific  $^{32}$ S-labeled probes for human lactoferrin. No detectable RNA hybridization is observed in the normal endometrium by *in situ* hybridization, even in the presence of immunodetectable protein, using equivalent hybridization conditions and development times as used for the adenocarcinomas. Consistent with the results obtained by *in situ* hybridization, long exposure times were required to demonstrate lactoferrin mRNA in normal endometrium by Northern blot using poly(A<sup>+</sup>) mRNA. This indicates that lactoferrin mRNA is present in normal

tissue but in very low levels, consistent with the very limited pattern of protein expression in normal endometrium. Equivalent RNA loading and quality for each specimen was demonstrated by ethidium bromide staining of the RNA gels (data not shown) and by probing for the housekeeping gene glyceraldehyde-3-phosphate dehydrogenase that does not fluctuate significantly with the metabolic state of the tissue. A representative Northern blot is shown in Fig. 5. *In situ* hybridization with several adenocarcinomas reveals that there is a direct correlation between the localization of lactoferrin mRNA and the immunostaining of expressed lactoferrin protein (Fig. 4). Lactoferrin mRNA is not associated with polymorphonuclear leukocytes by *in situ* hybridization in either normal or malignant tissue.

#### Western Blot Analysis

To confirm the specificity of the antisera that we used for immunohistochemistry, we performed Western blot analysis on proteins extracted from both normal and malignant endometrium which were separated by 8.5% SDS-polyacrylamide gel electrophoresis (Fig. 6). Immunoblotting identified a single broad protein band with a molecular weight between 70,000–80,000 in both normal and neoplastic endometrial tissue homogenates, consistent with the reported molecular weight of human lactoferrin. Supporting the immunocytochemical analysis, a representative immunoblot clearly demonstrates that the proportion of protein that is lactoferrin is markedly increased in the adenocarcinomas in comparison to the normal endometrium. The molecular weight of lactoferrin in the adenocarcinomas appears to have a slightly higher molecular weight than the predominant form in normal tissue.

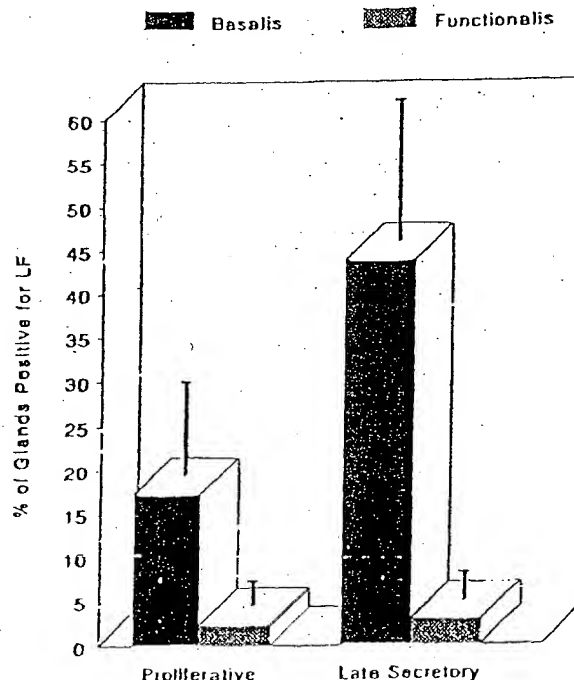


Fig. 2. Percentage of endometrial glands expressing lactoferrin protein by immunohistochemistry. Significantly more glands are positive in the region of the zona basalis than in the zona functionalis of the endometrium ( $P < 0.001$ ). Zona basalis, ■; zona functionalis, □. Although there is a trend towards more of the basalis glands expressing lactoferrin in the secretory phase (right ■ compared with the left ■), this was not statistically significant.

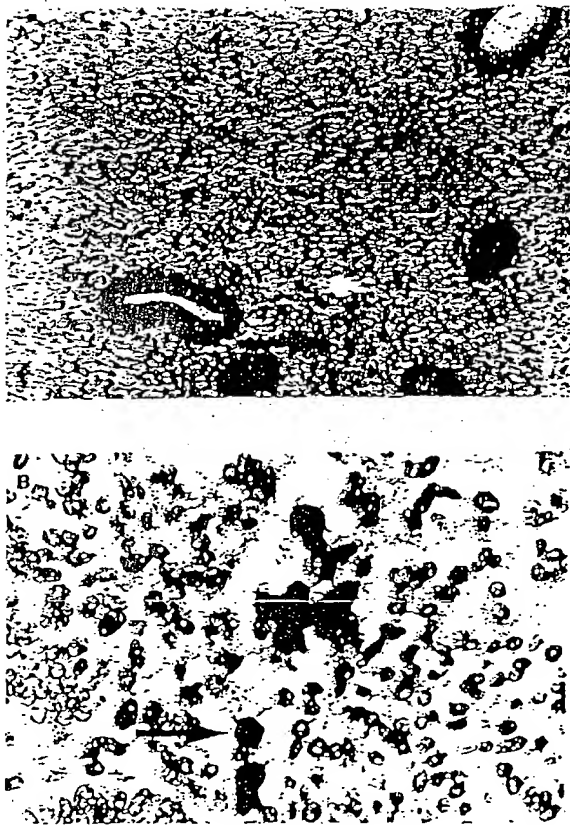


Fig. 3. Polymorphonuclear leukocytes (arrows) are scattered throughout the endometrium and stain intensely for lactoferrin. Lactoferrin is a known component of the secondary granules in polymorphonuclear leukocytes. The presence of a segmented nucleus and lactoferrin protein is an excellent method for identifying this group of inflammatory cells. A,  $\times 20$ . B,  $\times 60$ .

#### Correlation of Lactoferrin Expression with the Expression of PCNA, Ki-67, HER-2/*neu*, ER, and PR

In an attempt to characterize the phenotype of endometrial cells which express lactoferrin, we performed immunohistochemistry on serial sections for the Ki-67 antigen, PCNA, HER-2/*neu*, lactoferrin, ER, and PR. In normal tissue, Ki-67 and PCNA expression are cell cycle-specific markers of cell proliferation (16, 17). Upon analysis of normal cycling endometrium, no relationship between lactoferrin protein expression and ER, PR, or Ki-67 expression was observed. Similarly, in most of the adenocarcinomas evaluated, no relationship was noted between lactoferrin and PCNA protein expression. However, in one adenocarcinoma (Fig. 7), there was a clear inverse relationship seen between lactoferrin and PCNA localization, which was present throughout the entire tumor. Most dramatic, however, was a striking inverse correlation seen between lactoferrin and PR expression in 8 of 8 PR-positive uterine adenocarcinomas (Fig. 7). Two tumors negative for PR also did not express lactoferrin. Although an inverse correlation was also suggested with HER-2/*neu* and PR, the inverse correlation was more precise with lactoferrin in these tumors.

#### DISCUSSION

In the mouse uterus, lactoferrin is an estrogen-induced uterine secretory protein that is present throughout the epithelium (7), and it

is expressed concomitantly with epithelial cell proliferation. In contrast to lactoferrin's ubiquitous expression in the estrogenized mouse uterine epithelium, lactoferrin protein is limited to glandular epithelial cells in the basal regions of normal human endometrium and usually to glands that were directly adjacent to the myometrium (i.e., the deepest glands of the zona basalis). This regional localization of lactoferrin expression is not surprising in that other biochemical parameters have been reported to show site specificity in primate endometrium. These parameters include the proliferative index and the expression of the secretory component of IgA (18). Similar to our observations in the endometrium, lactoferrin is also expressed regionally in the mammary gland. In bovine breast tissue, lactoferrin is localized primarily to the basal alveolar cells (19); whereas in human breast tissue, the ductal epithelium appears to be the primary source of secreted lactoferrin during lactation (20).

Examining the endometrium on different days of the menstrual cycle demonstrates a trend towards increased lactoferrin expression during the luteal phase. Although this data is not statistically significant, the cyclic variation may be biologically relevant. Kim *et al.* (21) recently reported that the basal endometrial epithelial cells are unique because they proliferate during the postovulatory luteal phase. Interestingly, in the mouse uterus, there is a direct correlation between lactoferrin expression and epithelial cell proliferation (6). Therefore, lactoferrin may have a similar role in the human and mouse endometrium. If lactoferrin expression is cyclic, the ERE in the 5'-flanking promoter region of the human lactoferrin gene may be activated during the luteal phase (10, 11).

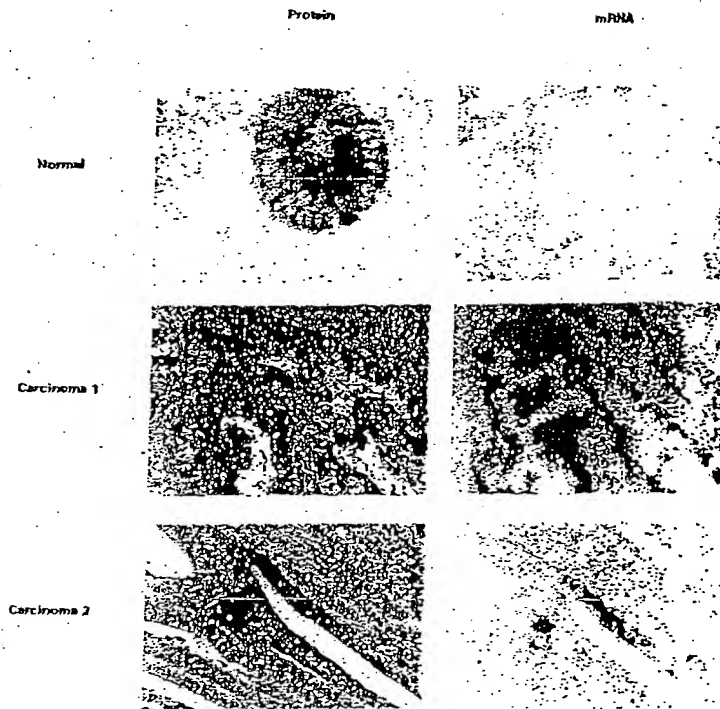
Another similarity between human and mouse uterine lactoferrin expression is the observation of two immunohistochemical staining patterns. In one pattern, the antisera binds primarily over the cytoplasm, and in the other, the nucleus is the primary site of localization. Although this could represent a fixation artifact, we have now observed this pattern in two species and under different fixation conditions. It has been demonstrated that signaling peptides, i.e., platelet-derived growth factor (22), Int-2 (23), and probasin (24), can be selectively directed into the nucleus, cytoplasm, or secreted. It is believed that this differential processing may allow proteins to have intracrine, autocrine, and paracrine roles, depending on the physiological state of the cell. The two localization patterns observed suggest that lactoferrin might have signaling sequences that direct the final destination of the mature peptide.

In normal endometrium, lactoferrin mRNA is present but in very low levels. Prolonged exposure times are needed to visualize the mRNA band with Northern analysis. Although with *in situ* hybridization the lactoferrin mRNA signal is easily seen in adenocarcinomas, we failed to localize the lactoferrin mRNA in normal endometrium using an equivalent exposure time. These low levels of mRNA in normal endometrium suggest that the synthesis and degradation of lactoferrin mRNA is more tightly regulated in normal tissue than in the adenocarcinomas. Because of the low levels of message in normal tissue, we are unable at this time to definitively conclude that lactoferrin mRNA is synthesized by the same epithelial cells which express the protein.

In endometrial adenocarcinomas, malignant transformation of the endometrium is associated with the up-regulation of lactoferrin mRNA and protein biosynthesis. The up-regulation at the RNA level is demonstrated by an increase in steady-state RNA levels using both *in situ* hybridization and Northern analysis techniques. In these cancers, we also observe an increase in the number of lactoferrin-positive cells, which express both the protein and mRNA. In this study, 8 of the 12 adenocarcinomas evaluated overexpress lactoferrin, compared with only 1 of 10 hyperplastic specimens. The form of lactoferrin protein extracted from endometrial adenocarcinomas appears to have

# LACTOFERRIN DYSREGULATION IN ENDOMETRIAL CANCER

Fig. 4. Colocalization of lactoferrin protein (left panels) and mRNA (right panels) in a normal proliferative endometrium and endometrial adenocarcinomas by performing immunohistochemistry and *in situ* RNA hybridization on serial sections. Dual analysis of protein and mRNA expression reveals that glands in normal endometrium do not have detectable mRNA, as measured by *in situ* hybridization (top panels;  $\times 40$ ), whereas analysis of the adenocarcinomas clearly demonstrates a direct correlation between protein and RNA expression for lactoferrin (middle and bottom panels;  $\times 10$ ). Note that lactoferrin protein and mRNA is distributed in a heterogeneous pattern in the epithelial cells of the adenocarcinomas. As is shown in Fig. 1, a heterogeneous staining pattern for lactoferrin protein is also seen frequently in normal endometrium.



a slightly higher molecular weight than the protein present in normal tissue. This could be due to alterations in the processing of the lactoferrin mRNA, protein, or glycosylation by the malignant cells. Alternatively, there could be minor differences between lactoferrin protein which is present in neutrophils and the form synthesized by uterine epithelial cells. We suggest two hypotheses to explain lactoferrin overexpression in endometrial adenocarcinomas. In the first hypothesis, lactoferrin biosynthesis is deregulated by the same processes that lead to the malignant transformation of endometrium. If this hypothesis is true then lactoferrin may be a useful marker for endometrial adenocarcinoma investigation, and further research is needed to determine whether lactoferrin plays a contributing role in the malignant transformation. A second hypothesis is that lactoferrin-positive human endometrial adenocarcinomas evolve from the clonal expansion of cells residing in the regenerative zone (zona basalis) of normal endometrium. It is interesting to speculate that lactoferrin expression in endometrial cancer may be linked to estrogen action in

some way, since proliferative disorders of human endometrium are linked to chronic estrogen exposure over several years and sequencing data suggests that the promoter for the human lactoferrin gene does contain an ERE.

Although the function of lactoferrin is unknown, a variety of biological roles have been proposed for lactoferrin which could link this protein to a role in cancer, including the regulation of DNA synthesis (25-29), modulation of the immune response (25, 30, 31), and iron transport (32). Some forms of lactoferrin are reported to have RNase activity (33, 34). Secreted RNases are involved in development, reproductive function, neoplasia, angiogenesis, and immune suppression. (35, 36) If angiogenesis and immunosuppression are components of lactoferrin RNase activity, these properties could promote tumor growth.

In the endometrial adenocarcinomas, we observed a heterogeneous expression pattern for lactoferrin, PCNA, Her-2/*neu*, ER, and PR. With regard to prognosis, patient survival is reportedly worse when

Table 1 Correlation of lactoferrin and PR expression in human endometrial adenocarcinomas

An inverse correlation indicates that lactoferrin and progesterone were not expressed in the same regions of the tumor by immunohistochemical analysis. PR was detected with the antisera KD68, and lactoferrin was detected by a specific polyclonal antisera.

Type of cancer	Age	Stage	Therapy	FIGO* grade	Lymph nodes	Myometrial invasion %	Lactoferrin expressed	Inverse correlation with PR
Endometrioid	37	1	None	1	0/12	0	No	No
Endometrioid	61	1	Estrogen	1-3	ND	60	Yes	Yes
Endometrioid	67	1	Estrogen	2-5	ND	60	Yes	Yes
Endometrioid, squamous differentiation	39	1a	None	2	ND	0	Yes	Yes
Endometrioid	69	1b	None	1	ND	29	No	No
Endometrioid	57	1b	None	1	ND	36	Yes	Yes
Endometrioid	72	1b	None	2-3	0/15	5	Yes	Yes
Endometrioid	67	1c	Estrogen	1	ND	50	Yes	Yes
Endometrioid, squamous differentiation	62	2a	None	2	ND	50	No	No
Endometrioid	78	2a	None	2	0/17	5	Yes	Yes
Endometrioid, squamous differentiation	63	3c	None	3	1/27	50	No	No
Endometrioid	54	4b	None	2-3	ND	75	Yes	Yes

\* FIGO, Federation Internationale des Gynecologues et Obstetristes; ND, not done.

1 2 3 4

Fig. 5. Northern analysis of lactoferrin mRNA expression confirms that endometrial adenocarcinomas (Lanes 3 and 4) significantly overexpress the 2.5-kilobase transcript of human lactoferrin in comparison to normal endometrium (Lanes 1 and 2). The Northern hybridization data supports the *in situ* RNA results and confirms that lactoferrin RNA expression is dysregulated in uterine adenocarcinomas. Normal uterine tissues appear to contain low steady-state RNA levels of lactoferrin, reflecting a controlled pattern of protein expression. Equivalent RNA loading and quality for each specimen was demonstrated by ethidium bromide staining of the RNA gels (data not shown) and by probing for a housekeeping gene (glyceraldehyde-3-phosphate dehydrogenase) that does not fluctuate significantly with the metabolic state of the tissue.

endometrial adenocarcinomas lose sex steroid receptors (37, 38), have a higher proliferative index (39), and demonstrate DNA aneuploidy. During the tumor progression of endometrial adenocarcinomas, it appears that the loss of steroid hormone receptors occurs earlier than either the increase in proliferation rate or the development of DNA aneuploidy (40). In our study, we note a striking inverse correlation between the expression of lactoferrin and PR in the endometrial adenocarcinomas. An inverse relationship also has been described for HER-2/*neu* and PR in endometrial adenocarcinomas that correlates with patient prognosis. Furthermore, in cancers of the human endometrium, ovary and breast Her-2/*neu* expression has been associated with advanced disease and poor survival (41-43). HER-2/*neu* is an oncogene that shares sequence homology with the epidermal growth factor receptor and is speculated to contribute to aberrant growth. Of note is that lactoferrin biosynthesis in the mouse uterus is associated with the expression of the epidermal growth factor. Like HER-2/*neu*, the epidermal growth factor receptor is also frequently overexpressed in PR-negative cells of endometrial adenocarcinomas (44). The amplification of growth factor receptor expression in PR-negative endometrial adenocarcinomas may be associated with the acquisition of growth autonomy and hormone independence, which may contribute to the poorer prognosis of PR-negative endometrial carcinomas (45). Some endometrial adenocarcinomas, including recurrent tumors, can be treated successfully with progesterone therapy (46, 47). Although the significance of the inverse relationship between lactoferrin and PR expression is not known, we speculate that the PR-negative cells do not undergo the normal growth inhibition and secretory differentiation normally associated with progesterone action.

A survey of human tissues reveals that lactoferrin is expressed by most normal mammalian exocrine glands and may be a prognostic marker in tumors (20). Lactoferrin is found in normal ductile breast epithelium and in primary breast carcinomas. In breast tumors, there is an inverse correlation between lactoferrin and ERs (20). Notably, lactoferrin expression in breast cancer may fall into the same category

Fig. 6. Western blot analysis using an antiserum specific for human lactoferrin was performed on proteins extracted from normal and malignant endometrium and separated by SDS-polyacrylamide gel electrophoresis. A broad protein band with a molecular weight between 70,000 and 80,000 is detected in both normal and neoplastic endometrial tissue homogenates, consistent with the reported molecular weight of human lactoferrin (A). The most significant observation from the immunoblotting studies is that lactoferrin protein is markedly elevated in the adenocarcinomas, in comparison to normal endometrium, which supports the immunocytochemical analysis that demonstrates a greater number of cells positive for lactoferrin protein in the uterine tumors. B, the relative amount of protein loaded in each lane by Coomassie blue staining.

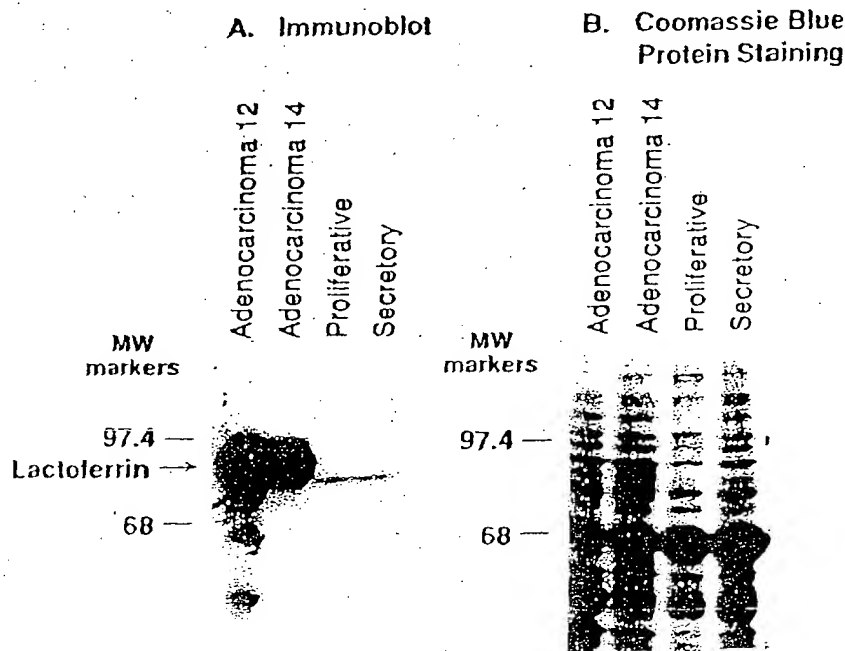
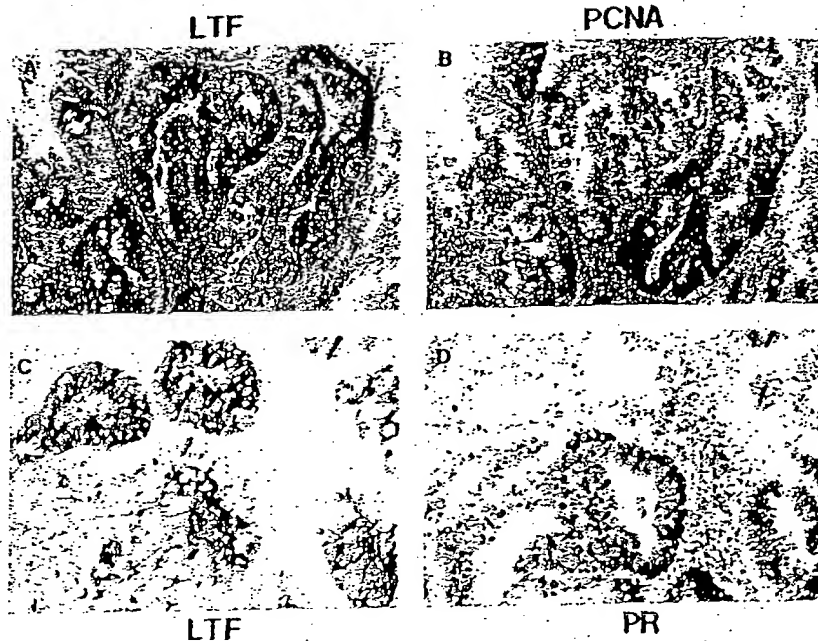




Fig. 7. Correlation of lactoferrin protein expression with the PCNA, a marker of proliferation, and PR as measured by immunohistochemistry performed on serial sections: X 20. In most cases, no relationship between lactoferrin and PCNA expression is found in either normal or malignant endometrium (data not shown). However, in 1 of 12 adenocarcinomas, an inverse correlation is seen between lactoferrin (A) and PCNA (B) localization. This striking pattern was consistent throughout the entire tumor, suggesting the possibility of cell cycle regulation of lactoferrin expression in this adenocarcinoma. The bottom panels exhibit an inverse correlation between lactoferrin (C) and PR expression (D), which was seen in all eight endometrial adenocarcinomas which expressed PR.



as the other markers for ER-negative tumors, such as amplification of EGF receptor, HER-2/*neu*, and transforming growth factor  $\alpha$  expression, which are associated with poor prognosis. In gastric carcinomas, lactoferrin expression is associated with transformation of specific cell types including intestinal-type carcinomas, adenomas, and incomplete intestinal metaplasias (28). Although this is complete speculation at this time, perhaps lactoferrin overexpression in the various malignancies may complement the actions of the growth factor pathway molecules and contribute to the autonomous growth of these tumors.

In conclusion, our studies reveal that lactoferrin is associated with a unique population of epithelial cells in the zona basalis and that lactoferrin overexpression may be associated with malignant transformation of the human endometrium. Further studies are needed to elucidate the role of lactoferrin in normal and pathological endometrial physiology.

#### ACKNOWLEDGMENTS

We gratefully acknowledge Dr. Geoffrey Greene from the Ben May Institute at the University of Chicago for providing the antisera to the PR (KD68) and the collaborative and collegial relationship between the intramural program at the National Institute of Environmental Health Sciences and the Department of Obstetrics and Gynecology at Duke University Medical Center that made this research possible.

#### REFERENCES

- Pokras, R., and Hulsopel, V. Hysterectomies in the United States. In: National Health Survey, DHHS Publication PHS 88-1753, Series 13, #92. pp. 1-26, 1987.
- Pollard, J. Regulation of polypeptide growth factor synthesis and growth factor-related gene expression in the rat and mouse uterus before and after implantation. *J. Reprod. Fertil.* 48: 721-731, 1990.
- Marian, L. Estrogens, anti-estrogens and the regulation of cell proliferation in the female reproductive tract in vivo. In: J. McLachlan (ed.), *Estrogens in the Environment*, pp. 105-121. New York: Elsevier North Holland, Inc., 1979.
- Marian, L., and Finn, C. A. Hormonal regulation of cell division in epithelial and connective tissues of the mouse uterus. *J. Endocrinol.* 41: 365-371, 1968.
- Green, M. R., and Pastewka, J. V. Lactoferrin is a marker for prolactin response in mouse mammary explants. *Endocrinology*, 103: 1510-1513, 1978.
- Walmer, D. K., Wrons, M. A., Hughes, C. L., and Nelson, K. G. Lactoferrin expression in the mouse reproductive tract during the natural estrous cycle: correlation with circulating estradiol and progesterone. *Endocrinology*, 131: 1458-1466, 1992.
- Teng, C. T., Pentecost, B. T., Chen, Y. H., Newbold, R. R., Eddy, E. M., and McLachlan, J. A. Lactoferrin gene expression in the mouse uterus and mammary gland. *Endocrinology*, 124: 992-999, 1989.
- McMaster, M. T., Teng, C. T., Dey, S. K., and Andrews, G. K. Lactoferrin in the mouse uterus: analyses of the preimplantation period and regulation by ovarian steroids. *Mol. Endocrinol.* 6: 101-111, 1992.
- Teng, C. T., Walker, M. P., Bhattacharya, S. N., Klapper, D. G., DiAugustine, R. P., and McLachlan, J. A. Purification and properties of an oestrogen-stimulated mouse uterine glycoprotein (approx. 70 kDa). *Biochem. J.* 240: 413-422, 1986.
- Rey, M. W., Woloshuk, S. L., deBoer, H. A., and Pieper, F. R. Complete nucleotide sequence of human mammary gland lactoferrin. *Nucleic Acids Res.* 18: 5286, 1990.
- Teng, C. T., Lin, Y., Yang, N., Walmer, D. K., and Panella, T. Differential molecular mechanism of the estrogen action that regulates lactoferrin gene in human and mouse. *Mol. Endocrinol.* 6: 1969-1981, 1992.
- Noyes, R. W., Henig, A. T., and Rock, J. Dating the endometrial biopsy. *Fertil. Steril.* 1: 3-25, 1950.
- Silverberg, S. G., and Kurman, R. J. Tumor classification. In: S. G. Silverberg and R. T. Kurman (eds.), *Tumors of the Uterine Corpus and Gestational Trophoblastic Disease*, pp. 15. Washington, DC: Armed Forces Institute of Pathology, 1992.
- Announcement, FIGO Stages: 1988 Revision. *Gynecol. Oncol.* 35: 125-126, 1989.
- Sambrook, J., Fritsch, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, 1989.
- Galand, P., and Degraef, C. Cyclin/PCNA immunostaining as an alternative to thymidine pulse labelling for marking S phase cells in paraffin sections from animal and human tissues. *Cell Tissue Res.* 22: 383-392, 1989.
- Kokuguchi, S., Hayase, R., and Sekiba, K. Proliferative activity in normal endometrium and endometrial carcinoma measured by immunohistochemistry using Ki-67 and anti-DNA polymerase  $\alpha$  antibody, and by flow cytometry. *Acta Medica Okayama*, 46: 115-121, 1992.
- Suzuki, M., Ogawa, M., Tamada, T., Nagata, H., and Watanabe, K. Immunohistochemical localization of secretory component and IgA in the human endometrium in relation to menstrual cycle. *Acta Histochem. Cytochem.* 17: 223-229, 1984.
- Huiley, W. L., and Rejman, J. J. Bovine lactoferrin in involuting mammary tissue. *Cell Biology Int.* 17: 283-289, 1993.
- Campbell, T., Skilton, R. A., Coombes, R. C., Shousha, S., Graham, M. D., and Luqmani, Y. A. Isolation of a lactoferrin cDNA clone and its expression in human breast cancer. *Br. J. Cancer*, 65: 19-26, 1992.
- Kim, S., Goodman, A. L., Williams, R. F., Hodgen, G. D., Hsu, J. G., and Chwalisz, K. Progesterone induces cyclic renewal in the "basalis" layer of primate endometrium: a possible site of action in anti-progesterone-induced endometrial regression. The American Fertility Society Annual Meeting Program, O-065, S32-S35, 1994. (Abstract)
- Lee, B. A., Maher, D. W., Hannink, M., and Donoghue, D. J. Identification of a signal for nuclear targeting in platelet-derived growth factor-related molecules. *Mol. Cell Biol.* 7: 3527-3537, 1987.



## LACTOFERRIN DYSREGULATION IN ENDOMETRIAL CANCER

23. Acland, P., Dixon, M., Peters, G., and Dickson, C. Subcellular fate of the int-2 oncoprotein is determined by choice of initiation codon. *Nature (Lond.)*, 343: 662-665, 1990.
24. Spence, A., Sheppard, P., Davis, J., Maton, Y., Nishi, N., McKeehan, W., Dodd, J., and Natusik, R. Regulation of a bifunctional mRNA results in synthesis of secreted and nuclear proteins. *Proc. Natl. Acad. Sci. USA*, 86: 7843-7847, 1989.
25. Gentile, P., and Broxmeyer, H. E. Interleukin-6 ablates the accessory cell-mediated suppressive effects of lactoferrin on human hematopoietic progenitor cell proliferation *in vitro*. *Ann. NY Acad. Sci.*, 628: 74-83, 1991.
26. Rejman, J. J., Oliver, S. P., Muenchen, R. A., and Turner, J. D. Proliferation of the MAC-T bovine mammary epithelial cell line in the presence of mammary secretion whey proteins. *Cell Biol. Int. Rep.*, 16: 993-1001, 1992.
27. Broxmeyer, H. E. Suppressor cytokines and regulation of myelopoiesis. Biology and possible clinical uses. *Am. J. Pediatr. Hematol. Oncol.*, 14: 22-30, 1992.
28. Tuccari, G., Barresi, G., Arena, F., and Inferrera, C. Immunocytochemical detection of lactoferrin in human gastric carcinomas and adenomas. *Arch. Pathol. Lab. Med.*, 113: 912-915, 1989.
29. Lognani, Y. A., Campbell, T. A., Bennett, C., Coombes, R. C., and Paterson, I. M. Expression of lactoferrin in human stomach. *Int. J. Cancer*, 49: 684-687, 1991.
30. Djcha, A., and Brock, J. H. Effect of transferrin, lactoferrin, and chelated iron on human T-lymphocytes. *Br. J. Haematol.*, 80: 235-241, 1992.
31. Broxmeyer, H. E., Maniel, C., Gentile, P., Sivastava, C., Miyazawa, K., Zucali, J. R., Rado, T. A., Levi, S., and Arosio, P. Actions of H-subunit ferritin and lactoferrin as suppressor molecules of myelopoiesis *in vitro* and *in vivo*. *Curr. Stud. Hematol. Blood Transfus.*, 178-181, 1991.
32. Sanchez, L., Calvo, M., and Brock, J. H. Biological role of lactoferrin. *Arch. Dis. Child.*, 67: 657-661, 1992.
33. Furmanski, P., and Li, Z. P. Multiple forms of lactoferrin in normal and leukemic human granulocytes. *Exp. Hematol.*, 18: 932-935, 1990.
34. Furmanski, P., Li, Z. P., Fortuna, M. B., Swamy, C. V., and Das, M. R. Multiple molecular forms of human lactoferrin: identification of a class of lactoferrins that possess ribonuclease activity and lack iron-binding capacity. *J. Exp. Med.*, 170: 415-429, 1989.
35. Fan, J. W., Snyder, D. J., Lobb, R. R., Alderman, E. M., Bethune, J. L., Riordan, J. F., and Valler, B. L. Isolation and characterization of angiogenin, an angiogenic protein from human carcinoma cells. *Biochemistry*, 24: 5480-5486, 1985.
36. D'Alessio, G., Di Donato, A., Parente, A., and Piccoli, R. Seminal RNase: a unique member of the ribonuclease superfamily. *Trends Biochem. Sci.*, 16: 104-106, 1991.
37. Mayer, T. K., and Mooney, R. A. Laboratory analyses for steroid hormone receptors, and their applications to clinical medicine. *Clin. Chim. Acta*, 172: 1-33, 1988.
38. Vihko, R., Alanko, A., Ikonen, V., and Kauppila, A. The predictive value of steroid hormone receptor analysis in breast, endometrial and ovarian cancer. *Med. Oncol. Tumor Pharmacother.*, 3: 197-210, 1986.
39. Rosenberg, P., Wingren, S., Simonsen, E., Stal, O., Risberg, B., and Nordenskjold, B. Flow cytometric measurements of DNA index and S-phase on paraffin-embedded early stage endometrial cancer: an important prognostic indicator. *Gynecol. Oncol.*, 33: 50-54, 1989.
40. Ponnonen, R., Mattila, J., Kuoppala, T., and Koivula, T. DNA ploidy, cell proliferation and steroid hormone receptors in endometrial hyperplasia and early adenocarcinoma. *J. Cancer Res. Clin. Oncol.*, 119: 426-429, 1993.
41. Slamon, D. J., Clark, G. M., Wong, S. G., Levin, W. J., Ulbrich, A., and McGuire, W. L. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/*neu* oncogene. *Science (Washington DC)*, 235: 177-182, 1987.
42. Slamon, D. J., Godolphin, W., Jones, L. A., et al. Studies of the HER-2/*neu* proto-oncogene in human breast and ovarian cancer. *Science (Washington DC)*, 244: 707-712, 1989.
43. Berchuck, A., Rodriguez, G., Kinney, R. B., Soper, J. T., Dodge, R. K., Clarke-Pearson, D. L., and Basl, R. C., Jr. Overexpression of HER-2/*neu* in endometrial cancer is associated with advanced stage disease. *Am. J. Obstet. Gynecol.*, 164: 15-21, 1991.
44. Bigsby, R. M., Li, A. X., Bomalaski, J., Suchman, F. B., Look, K. Y., and Soltan, G. P. Immunohistochemical study of HER-2/*neu*, epidermal growth factor receptor, and steroid receptor expression in normal and malignant endometrium. *Obstet. Gynecol.*, 79: 95-100, 1992.
45. Lindahl, B., Ferno, M., Gullberg, B., Norgren, A., and Willen, R. 5-year survival rate in endometrial carcinoma stage I-II related to steroid receptor concentration, degree of differentiation, age and myometrial invasion. *Anticancer Res.*, 12: 409-412, 1992.
46. Koborn, E. I. Gestagens and endometrial carcinoma. *Gynecol. Oncol.*, 4: 398-411, 1976.
47. Cressman, W. T. Clinical correlates of estrogen and progesterone binding protein in human endometrial carcinoma. *Obstet. Gynecol.*, 53: 363-376, 1980.

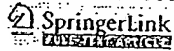
**Cell proliferation in human soft tissue tumors correlates with platelet-derived growth factor B chain expression: an immunohistochemical and in situ hybridization study.**

Wang J, Coltrera MD, Gown AM.

Department of Pathology, University of Washington, Seattle 98195.

The authors tested the hypothesis that the B chain of the platelet-derived growth factor (PDGF), a known connective tissue mitogen and growth factor, could be expressed by human soft tissue tumors, and that its expression could play a role in the control of cell proliferation in these tumors. Using a set of 56 soft tissue tumors, including benign tumors and all three grades of sarcomas, PDGF-B chain protein was localized using immunohistochemistry and PDGF-B mRNA was localized using in situ hybridization. The hypothesis that PDGF-B expression was related to cell proliferation was tested by simultaneously demonstrating the expression of the proliferating cell nuclear antigen in sequential tissue sections of the same tumors. Sixty and 82% of tumors had demonstrable PDGF-B mRNA and protein, respectively, with a strong correlation between their degrees of expression ( $P = 0.0001$ ). Among the sarcomas, a strong correlation between PDGF-B expression and increasing malignant tumor grade ( $P = 0.006$ ), and between PDGF-B expression and increasing proliferating cell nuclear antigen index ( $P = 0.01$ ) was found. All tumors were also demonstrated to express the beta receptor of PDGF via immunohistochemistry. These studies suggest that PDGF-B expression may be an important mediator of cell proliferation control, via an autocrine mechanism, in human soft tissue tumors and may correlate with clinical outcome in the sarcomas.

PMID: 7903911 [PubMed - indexed for MEDLINE]



**Expression of cadherins and catenins in paired tumor and non-neoplastic primary prostate cultures and corresponding prostatectomy specimens.**

Wang J, Krill D, Torbenson M, Wang Q, Bisceglia M, Stoner J, Thomas A, DeFlavia P, Dhir R, Becich MJ.

Department of Pathology, University of Pittsburgh Medical Center, PA, USA.

Cadherins are a family of transmembrane proteins that play a crucial role in cell differentiation, cell migration, and intercellular adhesion. Cadherins are associated with catenins through their highly conserved cytoplasmic domain. Down-regulation of E-cadherin protein has been shown in various human cancers. This study examined the expression of cadherins and associated catenins at the mRNA level. Paired tumor and nonneoplastic primary prostate cultures were obtained from surgical specimens. Quantitative multiplex fluorescence reverse transcriptase-polymerase chain reaction (QMF RT-PCR) and quantitative analysis were performed and correlated with immunostain results. Six of seven cases of neoplastic cultures showed moderately-to-markedly decreased levels of E-cadherin and P-cadherin mRNA. Similar losses of alpha-catenin and beta-catenin mRNA were also observed. The results of QMF RT-PCR showed good correlation with the results of immunohistochemical studies based on corresponding formalin-fixed sections. In conclusion, this paper presents a coordinated down-regulation in the expression of E-cadherin and associated catenins at the mRNA and protein level in most of the cases studied. This down-regulation may play an important role in the pathogenesis of prostate cancer.

PMID: 11127708 [PubMed - indexed for MEDLINE]

## ORIGINAL PAPER

Jianzhou Wang · Diane Krill · Michael Torbenson  
 Qi Wang · Michelle Bisceglia · Judy Stoner  
 Angela Thomas · Petrina DeFlavia · Rajiv Dhir  
 Michael J. Beich

## Expression of cadherins and catenins in paired tumor and non-neoplastic primary prostate cultures and corresponding prostatectomy specimens

Received: 6 March 2000 / Accepted: 23 May 2000

**Abstract** Cadherins are a family of transmembrane proteins that play a crucial role in cell differentiation, cell migration, and intercellular adhesion. Cadherins are associated with catenins through their highly conserved cytoplasmic domain. Down-regulation of E-cadherin protein has been shown in various human cancers. This study examined the expression of cadherins and associated catenins at the mRNA level. Paired tumor and non-neoplastic primary prostate cultures were obtained from surgical specimens. Quantitative multiplex fluorescence reverse transcriptase-polymerase chain reaction (QMF RT-PCR) and quantitative analysis were performed and correlated with immunostain results. Six of seven cases of neoplastic cultures showed moderately-to-markedly decreased levels of E-cadherin and P-cadherin mRNA. Similar losses of  $\alpha$ -catenin and  $\beta$ -catenin mRNA were also observed. The results of QMF RT-PCR showed good correlation with the results of immunohistochemical studies based on corresponding formalin-fixed sections. In conclusion, this paper presents a coordinated down-regulation in the expression of E-cadherin and associated catenins at the mRNA and protein level in most of the cases studied. This down-regulation may play an important role in the pathogenesis of prostate cancer.

**Key words** Prostate adenocarcinoma · Cadherin · Catenin · Adhesion molecules

J. Wang · D. Krill · M. Torbenson  
 Q. Wang · M. Bisceglia · J. Stoner · A. Thomas  
 P. DeFlavia · R. Dhir · M. J. Beich  
 Department of Pathology,  
 University of Pittsburgh Medical Center, Pittsburgh, PA  
 M. J. Beich (✉)  
 UPMC Shadyside,  
 Department of Pathology,  
 5230 Centre Avenue, Pittsburgh, PA 15232, USA  
 e-mail: beich@pitt.edu  
 Tel.: +1-412-6233941; Fax: +1-412-6232814

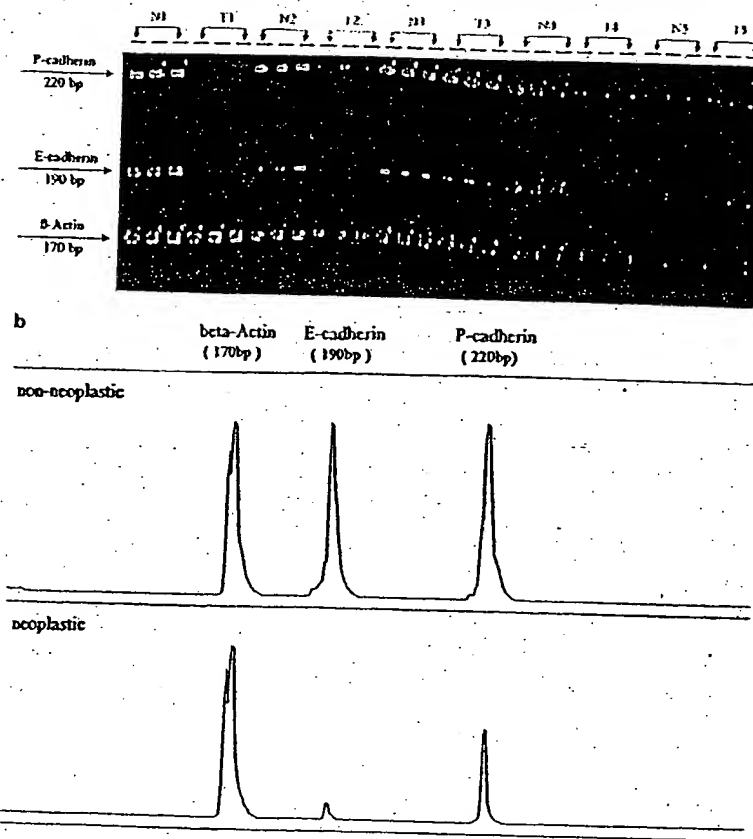
### Introduction

Prostate cancer is the most common malignant tumor and the second leading cause of cancer death in men. The clinical course of patients with prostate cancer varies widely, and different factors contribute to this marked clinical variability, including genetic background, hormonal environment, and the invasive potential of the tumor. Invasion and metastasis are the hallmarks of malignancy and have been closely linked to alterations in cell-to-cell adhesion, cell migration, and interactions with extracellular matrix components [22].

Cadherins are a family of transmembrane glycoproteins responsible for maintaining the integrity of tissue and are involved in cell differentiation, cell migration, and intercellular adhesion through a calcium-dependent mechanism characterized by homotypic adhesion [35–37]. Their highly conserved cytoplasmic domains associate with catenins, a group of intracellular proteins that mediate contact between the cadherins and the microfilaments of the cytoskeleton. Each cadherin subclass shows a unique tissue distribution: E-cadherin is predominantly expressed in epithelial cells and P-cadherin is restricted to decidua tissue and the basal or lower layers of stratified epithelium [30].

The accumulating evidence suggests a decrease or loss of function in E-cadherin and P-cadherin in several human carcinomas [3, 9, 23, 32]. Loss of heterozygosity (LOH) at chromosome 16 in the location of the E-cadherin gene is present in a high percentage of prostate cancers [6, 21, 26, 28]. Decreased expression of E-cadherin is seen in various human malignant tumor cell lines, and the level of decrease correlates with the invasive potential of the tumor cell lines [1, 9, 12, 24, 40, 42]. In addition, many, but not all, immunohistochemical studies using formalin-fixed, paraffin-embedded tissue have shown that the E-cadherin protein is decreased in prostate cancer and the decrease is correlated with tumor grade [4, 8, 13, 27, 38]. Other studies have shown decreased-to-absent P-cadherin levels, but variable E-cadherin levels [33].

**Fig. 1** A Genescan image of QMF RT-PCR of five paired non-neoplastic and neoplastic primary prostate cultures. cDNA corresponding to 50 ng of total RNA was subjected to 21 cycles of PCR, and all reactions were done in triplicate. RNA of both E-cadherin and P-cadherin showed moderate-to-marked reduction in primary neoplastic cultures in four of five cases (Cases 1, 2, 4, and 5). B Representative automated sequencer traces (ABI 373A) from 21 cycles of QMF RT-PCR using cDNA from a paired non-neoplastic (top panel) and neoplastic (bottom panel) primary prostate cultures (Case 2), and copy number of PCR products ( $\alpha$ -actin, E-cadherin, and P-cadherin) was calculated using peak areas. Neoplastic culture showed markedly decreased expression of both E-cadherin and P-cadherin



citrate buffer (pH 6.0) for antigen retrieval. The avidin-biotin complex method for immunohistochemistry was utilized from Vector Laboratories (Burlingame, Calif., USA). The E-cadherin, P-cadherin,  $\alpha$ -catenin, and  $\beta$ -catenin antibodies, all mouse monoclonal antibodies, were purchased from Transduction Laboratories (Lexington, Ky., USA).

#### Evaluation of immunostaining

The percentage of positive cells, intensity of the staining, and cellular localization of the staining were examined by two independent observers using normal prostate epithelium from the same specimen as an internal control. The intensity of the signal was graded as strong, moderate, weak, and negative. The staining pattern of the tumor was compared with that of normal epithelium from the same specimens.

#### Results

##### Expression of E-cadherin and P-cadherin

Since many prostate cancers are known to be histologically heterogeneous, adjacent H&E-stained sections of the tissue fragments sent to culture were reviewed

to assess tissue homogeneity and to rule out the presence of other diseases. The paired primary cultures for this study were selected based on the following histologic features: (1) non-neoplastic tissue showed no cancerous foci or high-grade prostatic intraepithelial neoplasia (PIN); (2) neoplastic sections contained less than 5% non-cancerous epithelium. Seven out of 38 pairs of cultures met the selection criteria and were included in this study. Histologically, all seven cases were moderately differentiated adenocarcinomas, with Gleason scores ranging from 5 to 7 (median = 6). In six of these cases, adjacent sections of tumor contained no benign prostate epithelium, and in one case (specimen 4), there was less than 5% non-cancerous epithelium. Prostate stromal cells express HGF, whereas the epithelial cells express c-met, the receptor for HGF [17, 20]. The cases included in this study showed no detectable HGF expression after 21 cycles of QMF RT-PCR (data not shown). This indicates there was no significant stromal cell contamination in the current epithelial cultures.

QMF-PCR is an accurate method of measuring the relative levels of mRNA in small tissue samples [41]. In

$\alpha$ - and  $\beta$ -catenins. Moderate interspecimen variation was observed in the baseline expression of  $\alpha$ -catenin and  $\beta$ -catenin mRNA levels in the non-neoplastic cultures.

#### Immunohistochemical studies of cadherins and catenins

In benign prostate tissue, E-cadherin was, in all cases, uniformly localized to the membranes of luminal glandular epithelial cells, predominantly at cell-cell junctions (Fig. 3A). One case of prostate cancer showed complete negative staining for E-cadherin (Fig. 3B), and the remaining six cases demonstrated reduced immunostaining for E-cadherin, with 25–75% of cancer cells positive (Fig. 3C). The cancerous glands generally showed reduced signal intensity and an altered heterogeneous staining pattern, which included focal cytoplasmic staining and reduced membranous staining (Fig. 3C, Table 4).

Benign prostate tissue showed uniform basal cell immunoreactivity for P-cadherin, with principally cytoplasmic and focal membranous pattern staining. The

benign, glandular non-basal epithelial cells and stromal cells were negative for P-cadherin (Fig. 3D). P-cadherin immunoreactivity was completely absent in two cases of prostate cancer (Fig. 3E), and the remaining five cases showed variable focal positivity, which was predominantly cytoplasmic (Fig. 3F). This focal P-cadherin immunostaining positivity was confirmed by staining multiple sections and by using different monoclonal antibodies. In some cases, serial sections also appeared to show immunostaining for both E- and P-cadherins with the same neoplastic cells.

In all cases of benign prostate tissue,  $\alpha$ - and  $\beta$ -catenin protein expression showed strong homogeneous staining of the luminal glandular epithelium and the basal cells. In a pattern similar to that of normal E-cadherin, the  $\alpha$ - and  $\beta$ -catenins were localized predominately at luminal epithelial cell-cell borders (Fig. 2A, C). In all cases of prostate cancer, there was a mild-to-moderate reduction in staining for both catenins with 50–75% of cells positive, and the staining tended to be heterogeneous (Fig. 2B, D; Table 4).

Table 4 Immunohistochemical expression of cadherins and catenins in prostatectomy specimens corresponding to primary prostate culture

Case	E-cadherin	P-cadherin	$\alpha$ -catenin	$\beta$ -catenin
1	–	–	++	++
2	++	+	++	++
3	+++	++	+++	+++
4	+	–	++	+++
5	+++	+	+++	+++
6	++	+	+++	+++
7	++	+	++	+++

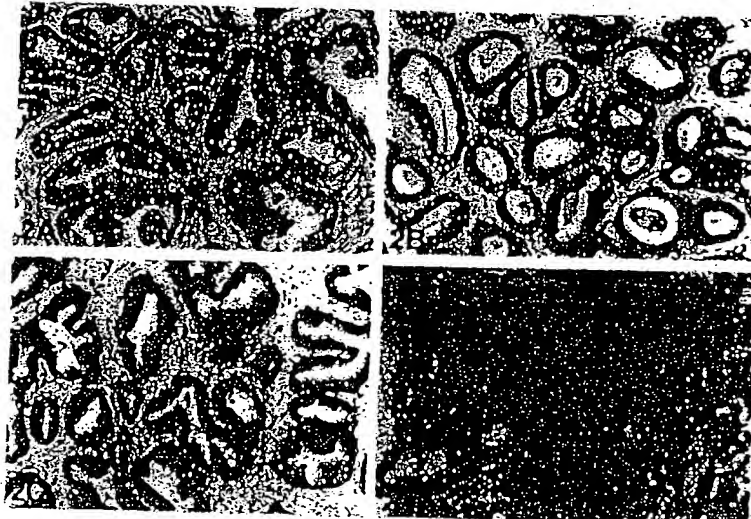
–, Negative; +, < 25% positive; ++, 25–50% positive; +++, 50–75% positive; + + + +, > 75% positive

#### Discussion

In this study, we observed a coordinated down-regulation of the expression in the genes involved in the cadherin and catenin mediated cell-cell pathway at the mRNA level. The protein levels, as demonstrated by the immunohistochemical studies on the corresponding tissue sections, were generally consistent with the mRNA data as well as with that reported in the literature [18, 33, 34, 38, 39].

E-cadherin showed the most consistent loss of expression at both the mRNA level and the protein level.

Fig. 2A–D Immunohistochemical staining of  $\alpha$ - and  $\beta$ -catenin in non-neoplastic prostate and in prostate adenocarcinoma. Original magnification  $\times 115$ . A, C Normal membranous expression of  $\alpha$ -catenin and  $\beta$ -catenin in non-neoplastic prostatic epithelium (Case 1). B Decreased immunostaining for  $\alpha$ -catenin in prostate adenocarcinoma (Case 1). D Decreased and heterogeneous immunostaining for  $\beta$ -catenin in prostate adenocarcinoma (Case 1)



appears to be derived by different combinations of a relatively few factors common to many types of tissues, and does not appear to be derived from transcription factors specific for each type of tissue. In our study, the degree of loss of expression of  $\alpha$ - and  $\beta$ -catenins at the mRNA level was correlated with reductions in the levels of E-cadherin expression. Although the exact mechanism of this coordinated down-regulation is not known, the coordinated pattern supports the hypothesis that loss or alteration of some regulatory factors occurs during prostate tumor progression. Possible mechanisms include transcriptional factor alterations or hypermethylation of the promoter region [14, 43]. Limited information is available on the transcriptional regulation of the catenins, but it is possible that similar mechanisms may play a role.

In summary, this is the first comparative study of the expression of the genes involved in the cadherin-mediated cell-cell adhesion pathway at the mRNA level using paired neoplastic and non-neoplastic primary cultures derived from prostatectomy specimens. Our results indicate that (1) there is a marked patient-to-patient variation in the normal levels of the cadherins and catenins; (2) mRNA levels of E-cadherin as well as catenins are significantly reduced in some prostate cancer primary cultures, and the reduction tends to be to the same degree in each tumor, suggesting a defect in a regulating mechanism common to all of these genes; (3) P-cadherin appears to be present at both the mRNA level and the protein level in some prostate cancers. This coordinated down-regulation of E-cadherin and catenin-mediated adhesion pathways may play a crucial role in tumor pathogenesis and metastasis.

## References

- Behrens J, Mareel M, Fiers W, van Roy F, Birchmeier W (1989) Dissecting tumor cell invasion: epithelial cells acquire invasive properties after the loss of uvomorulin-mediated cell-cell adhesion. *J Cell Biol* 108: 2435
- Behrens J, Lowrick O, Klein-Hitpass L, Birchmeier W (1991) The E-cadherin promoter: functional analysis of the GC-rich region and an epithelial cell-specific palindromic regulatory element. *Proc Natl Acad Sci USA* 88: 11,495
- Birchmeier W, Behrens J (1994) Cadherin expression in carcinomas: role in the formation of cell junctions and prevention of invasiveness. *Biochem Biophys Acta* 1198: 11
- Bussemakers MJ, van Moorselaar RJ, Girolodi LA, Ichikawa T, Isaacs JT, Takeichi M, Debruyne FM, Schalken JA (1992) Decreased expression of E-cadherin in the progression of rat prostate cancer. *Cancer Res* 52: 2916
- Bussemakers MJ, Girolodi LA, van Bokhoven A, Schalken JA (1994) Transcriptional regulation of the human E-cadherin gene in human prostate cell lines: characterization of the human E-cadherin promoter. *Biochem Biophys Res Commun* 203: 1284
- Carter BS, Ewing CM, Ward WS, Teiger BF, Alders TW, Scalken JA, Epstein JI, Isaacs WB (1990) Allelic loss of chromosomes 16q and 10q in human prostate cancer. *Proc Natl Acad Sci USA* 87: 8751
- Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Ann Biochem* 162: 156
- Cohen M, Griebing T, Ahaghotu C, Rokhlin O, Ross J (1997) Cellular adhesion molecules in urologic malignancies. *Am J Clin Pathol* 107: 56
- Doki Y, Shiozaki H, Tahara H, Inoue M, Oka H, Iihara K, Kadowaki T, Takeichi M, Mori T (1993) Correlation between E-cadherin expression and invasiveness in vitro in a human esophageal cancer cell line. *Cancer Res* 53: 3421
- Faraldo ML, Cano A (1993) The 5' flanking sequences of the mouse P-cadherin gene: homologies to 5' sequences of the E-cadherin gene and identification of a first 215 base-pair intron. *J Mol Biol* 231: 935
- Faraldo ML, Rodrigo I, Behrens J, Birchmeier W, Cano A (1997) Analysis of the E-cadherin and P-cadherin promoters in murine keratinocyte cell lines from different stages of mouse skin carcinogenesis. *Mol Carcinog* 20: 33
- Frixen UH, Behrens J, Sachs M, Eberle G, Voss B, Warda A, Lochner D, Birchmeier W (1991) E-cadherin-mediated cell-cell adhesion prevents invasiveness of human carcinoma cells. *J Cell Biol* 113: 173
- Girolodi LA, Schalken JA (1993) Decreased expression of the intracellular adhesion molecule E-cadherin in prostate cancer: biologic significance and clinical implications. *Cancer Metastasis Rev* 12: 29
- Graff JR, Herman JG, Lapidus RG, Chopra H, Xu R, Jarrard DF, Isaacs WB, Pitha PM, Davidson NE, Baylin SB (1995) E-cadherin expression is silenced by DNA hypermethylation in human breast and prostate carcinomas. *Cancer Res* 55: 5195
- Hatta M, Takeichi M (1994) Complex cell type-specific transcriptional regulation by the promoter and an intron of the mouse P-cadherin gene. *Dev Growth Differ* 36: 509
- Hinck L, Nathke IS, Papkoff J, Nelson WJ (1994) Dynamics of cadherin/catenin complex formation: novel protein interactions and pathways of complex assembly. *J Cell Biol* 125: 1327
- Humphrey PA, Zhu X, Zarnegar R, Swanson PD, Ratliff TL, Vollmer RT, Day ML (1995) Hepatocyte growth factor and its receptor (c-MET) in prostatic carcinoma. *Am J Pathol* 147: 386
- Jarrard DF, Paul R, Vanbokhoven A, Nguyen SH, Bova GS, Wheelock MJ, Johnson KR, Schalken J, Bussemakers M, Isaacs WB (1997) P-cadherin is a basal cell-specific epithelial marker that is not expressed in prostate cancer. *Clin Cancer Res* 3: 2121
- Kadowaki T, Shiozaki H, Inoue M, Tamura S, Oka H, Doki Y, Iihara K, Matsui S, Iwazawa T, Nagafuchi A (1994) E-cadherin and  $\alpha$ -catenin expression in human esophageal cancer. *Cancer Res* 54: 291
- Krill D, Shuman M, Thompson MT, Becich MJ, Strom SC (1997) A simple method for the isolation and culture of epithelial and stromal cells from benign and neoplastic prostates. *Urology* 49: 981
- Kunimi K, Bergerheim US, Larsson IL, Ekman P, Collins VP (1991) Allelotyping of prostate carcinoma. *Genomics* 11: 530
- Liotta LA (1986) Tumor invasion and metastasis: role of extracellular matrix. *Cancer Res* 46: 1
- Mayer B, Johnson JP, Leitz F, Jauch KW, Heiss MM, Schildberg FW, Birchmeier W, Funke I (1993) E-cadherin expression in primary and metastatic gastric cancer: down-regulation correlates with cellular dedifferentiation and glandular disintegration. *Cancer Res* 53: 1690
- Morton R, Ewing C, Nagafuchi A, Tsukita S, Isaacs W (1993) Reduction of E-cadherin levels and deletion of the  $\alpha$ -catenin gene in human prostate cancer cells. *Cancer Res* 53: 3585
- Murant SJ, Handley J, Stower M, Reid N, Cussenot O, Maitland NJ (1997) Coordinated changes in expression of cell-cell adhesion molecules in prostate cancer. *Eur J Cancer* 33: 263
- Natt E, Magen RE, Zimmer J, Mansouri A, Scherer G (1989) Regional assignment of the human foci for uvomorulin and chymotrypsinogen B with the help of two overlapping deletions on the long arm of chromosome 16. *Cytogenet Cell Genet* 50: 145
- Paul R, Ewing C, Jarrard D, Isaacs W (1997) The cadherin cell-cell adhesion pathway in prostate cancer progression. *Br J Urol* 79: 37

**Down-regulation of prostate-specific antigen expression by finasteride through inhibition of complex formation between androgen receptor and steroid receptor-binding consensus in the promoter of the PSA gene in LNCaP cells.**

Wang LG, Liu XM, Kreis W, Budman DR.

Department of Medicine, New York University, Manhasset 11030, USA.

As a specific competitive inhibitor of 5alpha-reductase, an intracellular enzyme that converts testosterone to dihydrotestosterone, finasteride is being extensively used for the treatment of benign prostatic hyperplasia and in experimental settings for prostate cancer. In this study, we showed that finasteride markedly inhibited prostate-specific antigen (PSA) secretion and expression. The promoter of the PSA gene contains several well-known cis-regulatory elements. Among them, steroid receptor-binding consensus (SRBC) has been identified as a functional androgen-responsive element. Our previous study showed that PSA was not only present in conditioned medium of the PSA-positive LNCaP cells but was also detectable in small amounts in PSA-negative cell lines, PC-3 and DU-145 (L. G. Wang et al., *Oncol. Rep.*, 3: 911-917, 1996). A strong correlation between binding of nuclear factors to SRBC and the level of PSA present in the conditioned medium and cell extracts was found in these three cell lines, whereas no such correlation with binding was obtained using Sp1 oligonucleotide as a probe. Binding of LNCaP cell nuclear proteins to SRBC was diminished when the cells were exposed to 25 microM finasteride, at which concentration 50% of both PSA mRNA and protein were inhibited. As a major component of DNA-protein complexes, the level of androgen receptor was dramatically decreased in the cells treated with finasteride. Our data indicate that inhibition of complex formation between SRBC and nuclear proteins due to the remarkable decrease in the level of androgen receptor plays a key role in the down-regulation of PSA gene expression by finasteride in LNCaP cells.

PMID: 9044850 [PubMed - indexed for MEDLINE]



### **Expression of calcyclin in human melanocytic lesions.**

**Weterman MA, van Muijen GN, Bloemers HP, Ruiter DJ.**

Department of Biochemistry, University of Nijmegen, The Netherlands.

When comparing two subsequent stages of melanocytic tumor progression we identified calcyclin as a new potential progression marker, the expression of which was correlated with metastatic behavior of various human melanoma cell lines in nude mice. In this study, we describe a good correlation between RNA and protein levels in the xenografts of these cell lines and extended these experiments to a panel of 120 routinely processed human melanocytic cutaneous lesions. Northern blot analysis demonstrated that calcyclin RNA expression was elevated in melanoma metastases as compared to several types of nevocellular nevi. Calcyclin staining using a specific polyclonal antiserum showed a more complex pattern. A stronger staining in a higher percentage of positive cells was observed in thick primary melanoma ( $\geq 1.5$  mm) as compared to thin primary melanoma ( $< 1.5$  mm). Calcyclin expression was also present in a higher percentage of cells showing a stronger staining in melanomas with higher Clark levels ( $> II$ ) corresponding to the vertical growth phase of primary melanomas. Protein expression in nevocellular nevi was confined to the dermal part and was highest in the lower parts of the dermis. Remarkably, dysplastic nevi (atypical moles), potential precursors of melanoma, did not show any expression at all, either in junctional or dermal parts. Confinement of the expression to the dermal part of nondysplastic nevi and primary melanomas may reflect interactions with the microenvironment of the reticular dermis that occurs with vertical growth.

PMID: 8261423 [PubMed - indexed for MEDLINE]



### Estrogen regulation of the cytochrome P450 3A subfamily in humans.

Williams ET, Leyk M, Wrighton SA, Davies PJ, Loose DS, Shipley GL, Strobel HW.

Department of Biochemistry, Medical School, University of Texas Health Science Center at Houston, 6431 Fannin, MSB 6.200, Houston, TX 77030, USA.

This study examines the possible role of estrogen in regulating the expression of the human CYP3A subfamily: CYP3A4, CYP3A5, CYP3A7, and CYP3A43. To accomplish this goal, mRNA was quantified from human livers and endometrial samples, and total CYP3A protein levels were evaluated by Western immunoblot analysis of the liver samples. The human endometrial samples were from premenopausal and postmenopausal women. The premenopausal endometrium was either in the proliferative or secretory phase, whereas for the postmenopausal endometrium samples, the women had been treated with either a placebo or estropipate, an estrogen substitute. After analyses, CYP3A4 mRNA was shown to have lower hepatic expression in females than in males. In the endometrium, CYP3A4 and CYP3A43 are down-regulated by estrogen, whereas CYP3A5 is expressed at higher levels during the secretory phase. CYP3A7 was not detected in the endometrium. In addition, the CYP3A subfamily showed increased mRNA expression in the liver as age increased. The expression levels of total CYP3A protein and total CYP3A mRNA showed good correlation. Despite apparent regulation of CYP3A4 mRNA expression by estrogen, the effects of estrogen may be overshadowed by additional regulators of gene expression.

PMID: 15282264 [PubMed - indexed for MEDLINE]

**Severely decreased MARCKS expression correlates with ras reversion but not with mitogenic responsiveness.**

Wojtaszek PA, Stumpo DJ, Blackshear PJ, Macara IG.

Department of Pathology, University of Vermont College of Medicine, Burlington 05405.

Phorbol ester-inducible phosphorylation of MARCKS, the '80-kDa' substrate of protein kinase C, was undetectable in several phenotypically dominant, non-transformed revertants independently derived from the ras-transformed cell line NIH3T3 DT-ras. Extremely low expression of MARCKS protein accounted for this apparent lack of phosphorylation. MARCKS-encoding mRNA levels were correspondingly decreased relative to normal and ras-transformed cells in all four ras revertant cell lines studied: C-11 and F-2, derived by 5-azacytidine treatment and selection with ouabain; CHP 9CJ, derived by ethylmethane sulfonate mutagenesis and selection with cis-hydroxy-L-proline; and 12-V3, derived by transfection with the human Krev-1 gene. However, re-expression of MARCKS after transfection of a cloned MARCKS cDNA into the C-11 ras revertant cells was not sufficient to induce retransformation. In fact, no significant difference in sensitivity to mitogenic stimulation by phorbol esters was observed among several cell lines expressing widely varying levels of MARCKS. This evidence argues against a direct role for MARCKS in mitogenic signaling. However, the strong correlation between attenuation of MARCKS expression and phenotypically dominant ras reversion suggests that a common negative regulatory mechanism might be responsible for both effects, presenting a potentially useful strategy for identifying factors involved in transducing the ras signal.

PMID: 8437859 [PubMed - indexed for MEDLINE]



Viewing options:

- Abstract
- Full text
- PDF (778KB)

Associated material:

- Readers' comments
- PubMed record

Related literature:

- Articles citing this article on BioMed Central on Google Scholar
- Other articles by authors
- Similar articles (PubMed)

Tools:

- E-mail to a friend
- Download references
- Post a comment

Key:

- ✉ E-mail
- ✉ Corresponding author

Research

## Prostate stem cell antigen (PSCA) expression in human prostate cancer tissues and its potential role in prostate carcinogenesis and progression of prostate cancer

Zhao Zhigang<sup>1</sup> ✉ and Shen Wenlv<sup>2</sup> ✉

<sup>1</sup>Department of Urology, Shantou University Medical College, Shantou, Guangdong, China

<sup>2</sup>Department of Urology, No 2. Affiliated Hospital of Shantou University Medical College, Shantou, Guangdong, China

World Journal of Surgical Oncology 2004, 2:13 doi:10.1186/1477-7819-2-13

The electronic version of this article is the complete one and can be found online at:  
<http://www.wjso.com/content/2/1/13>

Received 30 March 2004

Accepted 10 May 2004

Published 10 May 2004

© 2004 Zhigang and Wenlv; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

**Keywords:** Prostate, Neoplasm, Prostate stem cell antigen (PSCA)

Outline

### Abstract

### Background

Prostate stem cell antigen (PSCA) is a recently defined homologue of the Thy-1/Ly-6 family of glycosylphosphatidylinositol (GPI)-anchored cell surface antigens. The purpose of the present study was to examine the expression status of PSCA protein and mRNA in clinical specimens of human prostate cancer (Pca) and to validate it as a potential molecular target for diagnosis and treatment of Pca.

### Materials and Methods

Immunohistochemical (IHC) and *in situ* hybridization (ISH) analyses of PSCA expression were simultaneously performed on paraffin-embedded sections from 20 benign prostatic hyperplasia (BPH), 20 prostatic intraepithelial neoplasm (PIN) and 48 prostate cancer (Pca) tissues, including 9 androgen-independent prostate cancers. The level of PSCA expression was semiquantitatively scored by assessing both the percentage and intensity of PSCA-positive staining cells in the specimens. Then compared PSCA expression between BPH, PIN and Pca tissues and analysed the correlations

of PSCA expression level with pathological grade, clinical stage and progression to androgen-independence in Pca.

## Results

In BPH and low grade PIN, PSCA protein and mRNA staining were weak or negative and less intense and uniform than that seen in HGPIN and Pca. There were moderate to strong PSCA protein and mRNA expression in 8 of 11 (72.7%) HGPIN and in 40 of 48 (83.4%) Pca specimens examined by IHC and ISH analyses, with statistical significance compared with BPH (20%) and low grade PIN (22.2%) samples ( $p < 0.05$ , respectively). The expression level of PSCA increased with high Gleason grade, advanced stage and progression to androgen-independence ( $p < 0.05$ , respectively). In addition, IHC and ISH staining showed a high degree of correlation between PSCA protein and mRNA overexpression.

## Conclusions

Our data demonstrate that PSCA as a new cell surface marker is overexpressed by a majority of human Pca. PSCA expression correlates positively with adverse tumor characteristics, such as increasing pathological grade (poor cell differentiation), worsening clinical stage and androgen-independence, and speculatively with prostate carcinogenesis. PSCA protein overexpression results from upregulated transcription of PSCA mRNA. PSCA may have prognostic utility and may be a promising molecular target for diagnosis and treatment of Pca.

## Outline Introduction

Abstract  
Introduction  
Materials and methods  
Results  
Discussion  
Competing interests  
References

Prostate cancer (Pca) is the second leading cause of cancer-related death in American men and is becoming a common cancer increasing in China. Despite recently great progress in the diagnosis and management of localized disease, there continues to be a need for new diagnostic markers that can accurately discriminate between indolent and aggressive variants of Pca. There also continues to be a need for the identification and characterization of potential new therapeutic targets on Pca cells. Current diagnostic and therapeutic modalities for recurrent and metastatic Pca have been limited by a lack of specific target antigens of Pca.

Although a number of prostate-specific genes have been identified (i.e. prostate specific antigen, prostatic acid phosphatase, glandular kallikrein 2), the majority of these are secreted proteins not ideally suited for many immunological strategies. So, the identification of new cell surface antigens is critical to the development of new diagnostic and therapeutic approaches to the management of Pca.

Reiter RE et al [1] reported the identification of prostate stem cell antigen (PSCA), a cell surface antigen that is predominantly prostate specific. The PSCA gene encodes a 123 amino acid glycoprotein, with 30% homology to stem cell antigen 2 (Sca-2). Like Sca-2, PSCA also belongs to a member of the Thy-1/Ly-6 family and is anchored by a glycosylphosphatidylinositol (GPI) linkage. mRNA *in situ* hybridization (ISH) localized PSCA expression in

normal prostate to the basal cell epithelium, the putative stem cell compartment of prostatic epithelium, suggesting that PSCA may be a marker of prostate stem/progenitor cells.

In order to examine the status of PSCA protein and mRNA expression in human Pca and validate it as a potential diagnostic and therapeutic target for Pca, we used immunohistochemistry (IHC) and *in situ* hybridization (ISH) simultaneously; and conducted PSCA protein and mRNA expression analyses in paraffin-embedded tissue specimens of benign prostatic hyperplasia (BPH, n = 20), prostate intraepithelial neoplasm (PIN, n = 20) and prostate cancer (Pca, n = 48). Furthermore, we evaluated the possible correlation of PSCA expression level with Pca tumorigenesis, grade, stage and progression to androgen-independence.

## Outline Materials and methods

### Tissue samples

All of the clinical tissue specimens studied herein were obtained from 80 patients of 57–84 years old by prostatectomy, transurethral resection of prostate (TURP) or biopsies. The patients were classified as 20 cases of BPH, 20 cases of PIN, 40 cases of primary Pca, including 9 patients with recurrent Pca and a history of androgen ablation therapy (orchiectomy and/or hormonal therapy), who were referred to as androgen-independent prostate cancers. Eight specimens were harvested from these androgen-independent Pca patients prior to androgen ablation treatment. Each tissue sample was cut into two parts, one was fixed in 10% formalin for IHC and the other treated with 4% paraformaldehyde/0.1 M PBS PH 7.4 in 0.1% DEPC for 1 h for ISH analysis, and then embedded in paraffin. All paraffin blocks examined were then cut into 5  $\mu$ m sections and mounted on the glass slides specific for IHC and ISH respectively in the usual fashion. H&E-stained section of each Pca was evaluated and assigned a Gleason score by the experienced urological pathologist at our institution based on the criteria of Gleason score [2]. The Gleason sums are summarized in Table 1. Clinical staging was performed according to Jewett-whitmore-prout staging system, as shown in Table 2. In the category of PIN, we graded the specimens into two groups, i.e. low grade PIN (grade I – II) and high grade PIN (HGPIN, grade III) on the basis of literatures [3,4].

### Immunohistochemical (IHC) analysis

Briefly, tissue sections were deparaffinized, dehydrated, and subjected to microwaving in 10 mmol/L citrate buffer, PH 6.0 (Boshide, Wuhan, China) in a 900 W oven for 5 min to induce epitope retrieval. Slides were allowed to cool at room temperature for 30 min. A primary mouse antibody specific to human PSCA (Boshide, Wuhan, China) with a 1:100 dilution was applied to incubate with the slides at room temperature for 2 h. Labeling was detected by sequentially adding biotinylated secondary antibodies and streptavidin-peroxidase, and localized using 3,3'-diaminobenzidine reaction. Sections were then counterstained with hematoxylin. Substitution of the primary antibody with phosphate-buffered-saline (PBS) served as a negative-staining control.

## Outline

Abstract  
Introduction  
Materials and methods  
Results  
Discussion  
Competing interests  
References

## Tables

Table 1  
Correlation of PSCA expression with Gleason score

Table 2  
Correlation of PSCA expression with clinical stage

### mRNA *in situ* hybridization (ISH)

Five- $\mu$ m-thick tissue sections were deparaffinized and dehydrated, then digested in pepsin solution (4 mg/ml in 3% citric acid) for 20 min at 37.5°C, and further processed for ISH. Digoxigenin-labeled sense and antisense human PSCA RNA probes (obtained from Boshide, Wuhan, China) were hybridized to the sections at 48°C overnight. The posthybridization wash with a high stringency was performed sequentially at 37°C in 2  $\times$  standard saline citrate (SSC) for 10 min, in 0.5  $\times$  SSC for 15 min and in 0.2  $\times$  SSC for 30 min. The slides were then incubated to biotinylated mouse anti-digoxigenin antibody at 37.5°C for 1 h followed by washing in 1  $\times$  PBS for 20 min at room temperature, and then to streptavidin-peroxidase at 37.5°C for 20 min followed by washing in 1  $\times$  PBS for 15 min at room temperature. Subsequently, the slides were developed with diaminobenzidine and then counterstained with hematoxylin to localize the hybridization signals. Sections hybridized with the sense control probes routinely did not show any specific hybridization signal above background. All slides were hybridized with PBS to substitute for the probes as a negative control.

### Scoring methods

To determine the correlation between the results of PSCA immunostaining and mRNA *in situ* hybridization, the same scoring manners are taken in the present study for PSCA protein staining by IHC and PSCA mRNA staining by ISH. Each slide was read and scored by two independently experienced urological pathologists using Olympus BX-41 light microscopes. The evaluation was done in a blinded fashion. For each section, five areas of similar grade were analyzed semiquantitatively for the fraction of cells staining. Fifty percent of specimens were randomly chosen and rescored to determine the degree of interobserver and intraobserver concordance. There was greater than 95% intra- and interobserver agreement.

The intensity of PSCA expression evaluated microscopically was graded on a scale of 0 to 3+ with 3 being the highest expression observed (0, no staining; 1+, mildly intense; 2+, moderately intense; 3+, severely intense). The staining density was quantified as the percentage of cells staining positive for PSCA with the primary antibody or hybridization probe, as follows: 0 = no staining; 1 = positive staining in <25% of the sample; 2 = positive staining in 25%–50% of the sample; 3 = positive staining in >50% of the sample. Intensity score (0 to 3+) was multiplied by the density score (0–3) to give an overall score of 0–9 [1,5]. In this way, we were able to differentiate specimens that may have had focal areas of increased staining from those that had diffuse areas of increased staining [6]. The overall score for each specimen was then categorically assigned to one of the following groups: 0 score, negative expression; 1–2 scores, weak expression; 3–6 scores, moderate expression; 9 score, strong expression.

### Statistical analysis

Intensity and density of PSCA protein and mRNA expression in BPH, PIN and Pca tissues were compared using the Chi-square and Student's *t*-test. Univariate associations between PSCA expression and Gleason score, clinical stage and progression to androgen-independence were calculated using Fisher's Exact Test. For all analyses,  $p < 0.05$  was considered statistically significant.

## Outline Results

Abstract  
Introduction  
Materials and methods  
Results  
Discussion  
Competing interests  
References

## Figures



Figure 1  
Representatives of PSCA  
IHC and ISH staining in Pca  
(A. IHC staining, B. ISH  
staining,  $\times 200$   
magnification)

## PSCA expression in BPH

In general, PSCA protein and mRNA were expressed weakly in individual samples of BPH. Some areas of prostate expressed weak levels (composite score 1–2), whereas other areas were completely negative (composite score 0). Four cases (20%) of BPH had moderate expression of PSCA protein and mRNA (composite score 4–6) by IHC and ISH. In 2/20 (10%) BPH specimens, PSCA mRNA expression was moderate (composite score 3–6), but PSCA protein expression was weak (composite score 2) in one and negative (composite score 0) in the other. PSCA expression was localized to the basal and secretory epithelial cells, and prostatic stroma was almost negative staining for PSCA protein and mRNA in all cases examined.

## PSCA expression in PIN

In this study, we detected weak or negative expression of PSCA protein and mRNA ( $\leq 2$  scores) in 7 of 9 (77.8%) low grade PIN and in 2 of 11 (18.2%) HGPIN, and moderate expression (3–6 scores) in the rest 2 low grade PIN and 5 of 11 (45.5%) HGPIN. One HGPIN with moderate PSCA mRNA expression (6 score) was found weak staining for PSCA protein (2 score) by IHC. Strong PSCA protein and mRNA expression (9 score) were detected in the remaining 3 of 11 (27.3%) HGPIN. There was a statistically significant difference of PSCA protein and mRNA expression levels observed between HGPIN and BPH ( $p < 0.05$ ), but no statistical difference reached between low grade PIN and BPH ( $p > 0.05$ ).

## PSCA expression in Pca

In order to determine if PSCA protein and mRNA can be detected in prostate cancers and if PSCA expression levels are increased in malignant compared with benign glands, Forty-eight paraffin-embedded Pca specimens were analysed by IHC and ISH. It was shown that 19 of 48 (39.6%) Pca samples stained very strongly for PSCA protein and mRNA with a score of 9 and another 21 (43.8%) specimens displayed moderate staining with scores of 4–6 (Figure 1). In addition, 4 specimens with moderate to strong PSCA mRNA expression (scores of 4–9) had weak protein staining (a score of 2) by IHC analyses. Overall, Pca expressed a significantly higher level of PSCA protein and mRNA than any other specimen category in this study ( $p < 0.05$ , compared with BPH and PIN respectively). The result demonstrates that PSCA protein and mRNA are overexpressed by a majority of human Pca.

## Correlation of PSCA expression with Gleason score in Pca

Using the semi-quantitative scoring method as described in Materials and Methods, we compared the expression level of PSCA protein and mRNA with Gleason grade of Pca, as shown in Table 1. Prostate adenocarcinomas were graded by Gleason score as 2–4 scores = well-differentiation, 5–7 scores = moderate-differentiation and 8–10 scores = poor-differentiation [7]. Seventy-two percent of Gleason scores 8–10 prostate cancers had very



strong staining of PSCA compared to 21% with Gleason scores 5–7 and 17% with 2–4 respectively, demonstrating that poorly differentiated Pca had significantly stronger expression of PSCA protein and mRNA than moderately and well differentiated tumors ( $p < 0.05$ ). As depicted in Figure 1, IHC and ISH analyses showed that PSCA protein and mRNA expression in several cases of poorly differentiated Pca were particularly prominent, with more intense and uniform staining. The results indicate that PSCA expression increases significantly with higher tumor grade in human Pca.

### Correlation of PSCA expression with clinical stage in Pca

With regards to PSCA expression in every stage of Pca, we showed the results in Table 2. Seventy-five percent of locally advanced and node positive cancers (i.e. C-D stages) expressed statistically high levels of PSCA versus 32.5% that were organ confined (i.e. A-B stages) ( $p < 0.05$ ). The data demonstrate that PSCA expression increases significantly with advanced tumor stage in human Pca.

### Correlation of PSCA expression with androgen-independent progression of Pca

All 9 specimens of androgen-independent prostate cancers stained positive for PSCA protein and mRNA. Eight specimens were obtained from patients managed prior to androgen ablation therapy. Seven of eight (87.5%) of these androgen-independent prostate cancers were in the strongest staining category (score = 9), compared with three out of eight (37.5%) of patients with androgen-dependent cancers ( $p < 0.05$ ). The results demonstrate that PSCA expression increases significantly with progression to androgen-independence of human Pca.

It is evident from the results above that within a majority of human prostate cancers the level of PSCA protein and mRNA expression correlates significantly with increasing grade, worsening stage and progression to androgen-independence.

### Correlation of PSCA immunostaining and mRNA *in situ* hybridization

In all 88 specimens surveyed herein, we compared the results of PSCA IHC staining with mRNA ISH analysis. Positive staining areas and its intensity and density scores evaluated by IHC were identical to those seen by ISH in 79 of 88 (89.8%) specimens (18/20 BPH, 19/20 PIN and 42/48 Pca respectively). Importantly, 27/27 samples with PSCA mRNA composite scores of 0–2, 32/36 samples with scores of 3–6 and 22/24 samples with a score of 9 also had PSCA protein expression scores of 0–2, 3–6 and 9 respectively. However, in 5 samples with PSCA mRNA overall scores of 3–6 and in 2 with scores of 9 there were less or negative PSCA protein expression (i.e. scores of 0–4), suggesting that this may reflect posttranscriptional modification of PSCA or that the epitopes recognized by PSCA mAb may be obscured in some cancers. The data demonstrate that the results of PSCA immunostaining were consistent with those of mRNA ISH analysis, showing a high degree of correlation between PSCA protein and mRNA expression.

## Outline Discussion

Abstract  
Introduction  
Materials and methods  
Results  
Discussion  
Competing interests  
References

PSCA is homologous to a group of cell surface proteins that mark the earliest phase of hematopoietic development. PSCA mRNA expression is prostate-specific in normal male tissues and is highly up-regulated in both androgen-dependent and-independent Pca xenografts (LAPC-4 tumors). We hypothesize that PSCA may play a role in Pca tumorigenesis and progression, and may serve as a target for Pca diagnosis and treatment. In this study, IHC and ISH showed that in general there were weak or absent PSCA protein and mRNA expression in BPH and low grade PIN tissues. However, PSCA protein and mRNA are widely expressed in HGPIN, the putative precursor of invasive Pca, suggesting that up-regulation of PSCA is an early event in prostate carcinogenesis. Recently, Reiter RE et al [1], using ISH analysis, reported that 97 of 118 (82%) HGPIN specimens stained strongly positive for PSCA mRNA. A very similar finding was seen on mouse PSCA (mPSCA) expression in mouse HGPIN tissues by Tran C. P et al [8]. These data suggest that PSCA may be a new marker associated with transformation of prostate cells and tumorigenesis.

We observed that PSCA protein and mRNA are highly expressed in a large percentage of human prostate cancers, including advanced, poorly differentiated, androgen-independent and metastatic cases. Fluorescence-activated cell sorting and confocal/ immunofluorescent studies demonstrated cell surface expression of PSCA protein in Pca cells [9]. Our IHC expression analysis of PSCA shows not only cell surface but also apparent cytoplasmic staining of PSCA protein in Pca specimens (Figure 1). One possible explanation for this is that anti-PSCA antibody can recognize PSCA peptide precursors that reside in the cytoplasm. Also, it is possible that the positive staining that appears in the cytoplasm is actually from the overlying cell membrane [5]. These data seem to indicate that PSCA is a novel cell surface marker for human Pca.

Our results show that elevated level of PSCA expression correlates with high grade (i.e. poor differentiation), increased tumor stage and progression to androgen-independence of Pca. These findings support the original IHC analyses by Gu Z et al [9], who reported that PSCA protein expressed in 94% of primary Pca and the intensity of PSCA protein expression increased with tumor grade, stage and progression to androgen-independence. Our results also collaborate the recent work of Han KR et al [10], in which the significant association between high PSCA expression and adverse prognostic features such as high Gleason score, seminal vesicle invasion and capsular involvement in Pca was found. It is suggested that PSCA overexpression may be an adverse predictor for recurrence, clinical progression or survival of Pca. Hara H et al [11] used RT-PCR detection of PSA, PSMA and PSCA in 1 ml of peripheral blood to evaluate Pca patients with poor prognosis. The results showed that among 58 Pca patients, each PCR indicated the prognostic value in the hierarchy of PSCA>PSA>PSMA RT-PCR, and extraprostatic cases with positive PSCA PCR indicated lower disease-progression-free survival than those with negative PSCA PCR, demonstrating that PSCA can be used as a prognostic factor. Dubey P et al [12] reported that elevated numbers of PSCA + cells correlate positively with the onset and development of prostate carcinoma over a long time span in the prostates of the TRAMP and PTEN +/- models compared with its

normal prostates. Taken together with our present findings, in which PSCA is overexpressed from HGPIN to almost frank carcinoma, it is reasonable and possible to use increased PSCA expression level or increased numbers of PSCA-positive cells in the prostate samples as a prognostic marker to predict the potential onset of this cancer. These data raise the possibility that PSCA may have diagnostic utility or clinical prognostic value in human Pca.

The cause of PSCA overexpression in Pca is not known. One possible mechanism is that it may result from PSCA gene amplification. In humans, PSCA is located on chromosome 8q24.2 [1], which is often amplified in metastatic and recurrent Pca and considered to indicate a poor prognosis [13-15]. Interestingly, PSCA is in close proximity to the c-myc oncogene, which is amplified in >20% of recurrent and metastatic prostate cancers [16,17]. Reiter RE et al [18] reported that PSCA and MYC gene copy numbers were co-amplified in 25% of tumors (five out of twenty), demonstrating that PSCA overexpression is associated with PSCA and MYC coamplification in Pca. Gu Z et al [9] recently reported that in 102 specimens available to compare the results of PSCA immunostaining with their previous mRNA ISH analysis, 92 (90.2%) had identically positive areas of PSCA protein and mRNA expression. Taken together with our findings, in which we detected moderate to strong expression of PSCA protein and mRNA in 34 of 40 (85%) Pca specimens examined simultaneously by IHC and ISH analyses, it is demonstrated that PSCA protein and mRNA overexpressed in human Pca, and that the increased protein level of PSCA was resulted from the upregulated transcription of its mRNA.

At present, the regulation mechanisms of human PSCA expression and its biological function are yet to be elucidated. PSCA expression may be regulated by multiple factors [18]. Watabe T et al [19] reported that transcriptional control is a major component regulating PSCA expression levels. In addition, induction of PSCA expression may be regulated or mediated through cell-cell contact and protein kinase C (PKC) [20]. Homologues of PSCA have diverse activities, and have themselves been involved in carcinogenesis. Signalling through SCA-2 has been demonstrated to prevent apoptosis in immature thymocytes [21]. Thy-1 is involved in T cell activation and transduces signals through src-like tyrosine kinases [22]. Ly-6 genes have been implicated both in tumorigenesis and in cell-cell adhesion [23-25]. Cell-cell or cell-matrix interaction is critical for local tumor growth and spread to distal sites. From its restricted expression in basal cells of normal prostate and its homology to SCA-2, PSCA may play a role in stem/progenitor cell function, such as self-renewal (i.e. anti-apoptosis) and/or proliferation [1]. Taken together with the results in the present study, we speculate that PSCA may play a role in tumorigenesis and clinical progression of Pca through affecting cell transformation and proliferation. From our results, it is also suggested that PSCA as a new cell surface antigen may have a number of potential uses in the diagnosis, therapy and clinical prognosis of human Pca. PSCA overexpression in prostate biopsies could be used to identify patients at high risk to develop recurrent or metastatic disease, and to discriminate cancers from normal glands in prostatectomy samples. Similarly, the detection of PSCA-overexpressing cells in bone marrow or peripheral blood may identify and predict metastatic progression better than current assays, which identify only PSA-positive or PSMA-positive prostate cells.

In summary, we have shown in this study that PSCA protein and mRNA are maintained in expression from HGPIN through all stages of Pca in a majority of cases, which may be associated with prostate carcinogenesis and correlate positively with high tumor grade (poor cell differentiation), advanced stage and androgen-independent progression. PSCA protein overexpression is due to the upregulation of its mRNA transcription. The results suggest that PSCA may be a promising molecular marker for the clinical prognosis of human Pca and a valuable target for diagnosis and therapy of this tumor.

### Competing interests

None declared.

### Outline References

Abstract  
Introduction  
Materials and methods  
Results  
Discussion  
Competing interests  
References

1. Reiter RE, Gu Z, Watabe T, Thomas G, Szigeti K, David E, Wahl M, Nisitani S, Yamashiro J, Le Beau MM, Loda M, Witte ON: **Prostate stem cell antigen: a cell surface marker overexpressed in prostate cancer.** *Proc Natl Acad Sci USA* 1998, **95**:1735-1740. [PubMed Abstract] [Publisher Full Text]  
Return to citation in text: [1] [2] [3] [4] [5]
2. Gleason DF: **Histologic grading and clinical staging of prostatic carcinoma.** *In: Urologic Pathology: The Prostate (Edited by: Tannebaum M).* Philadelphia, Lea & Febiger 1977, 171-197.  
Return to citation in text: [1]
3. Brawer MK: **Prostatic intraepithelial neoplasia: a premalignant lesion.** *Hum Pathol* 1992, **23**:242-248. [PubMed Abstract]  
Return to citation in text: [1]
4. Amin MB, Ro JY, Ayala AC: **Prostatic intraepithelial neoplasia: relationship to adenocarcinoma of prostate.** *Pathol Annu* 1994, **29**:1-30. [PubMed Abstract]  
Return to citation in text: [1]
5. Amara N, Palapattu GS, Schrage M, Gu Z, Thomas GV, Dorey F, Said J, Reiter RE: **Prostate stem cell antigen is overexpressed in human transitional cell carcinoma.** *Cancer Res* 2001, **61**:4660-4665. [PubMed Abstract] [Publisher Full Text]  
Return to citation in text: [1] [2]
6. Hanas JS, Lerner NR, Lightfoot SA, Raczowski C, Kastens DJ, Brackett DJ, Postier RG: **Expression of the cyclin-dependent kinase inhibitor p21 (WAF1/CIP1) and p53 tumor suppressor in dysplastic progression and adenocarcinoma in Barrett esophagus.** *Cancer (Phila)* 1999, **86**:756-763. [Publisher Full Text]  
Return to citation in text: [1]
7. Egevad L, Gramfors T, Karlberg L: **Prognostic value of the Gleason score in prostate cancer.**


*BJU Int* 2002, **89**:538-542. [PubMed Abstract] [Publisher Full Text]  
Return to citation in text: [1]

8. Tran CP, Lin C, Yamashiro J, Reiter RE: **Prostate stem cell antigen is a marker of late intermediate prostate epithelial cells.**  
*Mol Cancer Res* 2002, **1**:113-121. [PubMed Abstract] [Publisher Full Text]  
Return to citation in text: [1]
9. Gu Z, Thomas G, Yamashiro J, Shintaku IP, Dorey F, Raitano A, Witte ON, Said JW, Loda M, Reiter RE: **Prostate stem cell antigen (PSCA) expression increases with high Gleason score, advanced stage and bone metastasis in prostate cancer.**  
*Oncogene* 2000, **19**:1288-1296. [PubMed Abstract] [Publisher Full Text]  
Return to citation in text: [1] [2] [3]
10. Han KR, Seligson DB, Liu X, Horvath S, Shintaku PJ, Thomas GV, Said JW, Reiter RE: **Prostate stem cell antigen expression is associated with gleason score, seminal vesicle invasion and capsular invasion in prostate cancer.**  
*J Urol* 2004, **171**:1117-1121. [PubMed Abstract] [Publisher Full Text]  
Return to citation in text: [1]
11. Hara H, Kasahara T, Kawasaki T, Bilim V, Obara K, Takahashi K, Tomita Y: **Reverse Transcription-Polymerase Chain Reaction Detection of Prostate-specific Antigen, Prostate-specific Membrane Antigen, and Prostate Stem Cell Antigen in One Milliliter of Peripheral Blood.**  
*Clin Cancer Res* 2002, **8**:1794-1799. [PubMed Abstract] [Publisher Full Text]  
Return to citation in text: [1]
12. Dubey P, Wu H, Reiter RE, Witte ON: **Alternative pathways to prostate carcinoma activate prostate stem cell antigen expression.**  
*Cancer Res* 2001, **61**:3256-3261. [PubMed Abstract] [Publisher Full Text]  
Return to citation in text: [1]
13. Visa korpi T, Kallioniemi AH, Syvanen AC, Hyytinen ER, Karhu R, Tammela T, Isola JJ, Kallioniemi OP: **Genetic changes in primary and recurrent prostate cancer by comparative genomic hybridization.**  
*Cancer Res* 1995, **55**:342-347. [PubMed Abstract]  
Return to citation in text: [1]
14. Sato K, Qian J, Slezak JM, Lieber MM, Bostwick DG, Bergstralh EJ, Jenkins RB: **Clinical significance of alterations of chromosome 8 in high-grade, advanced, nonmetastatic prostate carcinoma.**  
*J Natl Cancer Inst* 1999, **91**:1574-1580. [PubMed Abstract] [Publisher Full Text]  
Return to citation in text: [1]
15. Van Den Berg C, Guan XY, Von Hoff D, Jenkins R, Bittner J, Griffin C, Kallioniemi O, Visakorpi, McGill J, Herath J, Epstein J, Sarosdy M, Meltzer P, Trent J: **DNA sequence amplification in human prostate cancer identified by chromosome microdissection: potential prognostic implications.**  
*Clin Cancer Res* 1995, **1**:11-18. [PubMed Abstract]  
Return to citation in text: [1]
16. Jenkins RB, Qian J, Lieber MM, Bostwick DG: **Detection of c-myc oncogene amplification and chromosomal anomalies in metastatic prostatic carcinoma by fluorescence in situ hybridization.**  
*Cancer Res* 1997, **57**:524-531. [PubMed Abstract]  
Return to citation in text: [1]

17. Nupponen NN, Kakkola L, Koivisto P, Visakorpi T: **Genetic alterations in hormone-refractory recurrent prostate carcinomas.**  
*Am J Pathol* 1998, **153**:141-148. [PubMed Abstract] [Publisher Full Text]  
Return to citation in text: [1]
18. Reiter RE, Sato I, Thomas G, Qian J, Gu Z, Watabe T, Loda M, Jenkins RB: **Coamplification of prostate stem cell antigen (PSCA) and MYC in locally advanced prostate cancer.**  
*Genes Chromosomes Cancer* 2000, **27**:95-103. [PubMed Abstract] [Publisher Full Text]  
Return to citation in text: [1] [2]
19. Watabe T, Lin M, Donjacour AA, Cunha GR, Witte ON, Reiter RE: **Growth, regeneration, and tumorigenesis of the prostate activates the PSCA promoter.**  
*Proc Natl Acad Sci USA* 2002, **99**:401-406. [PubMed Abstract] [Publisher Full Text]  
Return to citation in text: [1]
20. Bahrenberg G, Brauers A, Joost HG, Jakse G: **PSCA expression is regulated by phorbol ester and cell adhesion in the bladder carcinoma cell line RT112.**  
*Cancer Lett* 2001, **168**:37-43. [PubMed Abstract] [Publisher Full Text]  
Return to citation in text: [1]
21. Noda S, Kosugi A, Saitoh S, Narumiya S, Hamaoka T: **Protection from anti-TCR/CD3-induced apoptosis in immature thymocytes by a signal through thymic shared antigen-1/stem cell antigen-2.**  
*J Exp Med* 1996, **183**:2355-2360. [PubMed Abstract]  
Return to citation in text: [1]
22. Thomas PM, Samelson LE: **The glycoposphatidylinositol-anchored Thy-1 molecule interacts with the p60fyn protein tyrosine kinase in T cells.**  
*J Biol Chem* 1992, **267**:12317-12322. [PubMed Abstract] [Publisher Full Text]  
Return to citation in text: [1]
23. Bamezai A, Rock KL: **Overexpressed Ly-6A.2 mediated cell-cell adhesion by binding a ligand expressed on lymphoid cells.**  
*Proc Natl Acad Sci USA* 1995, **92**:4294-4298. [PubMed Abstract] [Publisher Full Text]  
Return to citation in text: [1]
24. Katz BZ, Eshel R, Sagi-Assif O, Witz IP: **An association between high Ly-6A/E expression on tumor cells and a highly malignant phenotype.**  
*Int J Cancer* 1994, **59**:684-691. [PubMed Abstract]  
Return to citation in text: [1]
25. Brakenhoff RH, Gerretsen M, Knippels EM, van Dijk M, van Essen H, Weghuis DO, Sinke RJ, Snow GB, van Dongen GA: **The human E48 antigen, highly homologous to the murine Ly-6 antigen ThB, is a GPI-anchored molecule apparently involved in keratinocyte cell-cell adhesion.**  
*J Cell Biol* 1995, **129**:1677-1689. [PubMed Abstract] [Publisher Full Text]  
Return to citation in text: [1]

Have something to say? Post a comment on this article!



Published by  BioMed Central

© 1999-2006 BioMed Central Ltd unless otherwise stated < [info@biomedcentral.com](mailto:info@biomedcentral.com) > Terms and conditions

**Expression of superoxide dismutases, catalase, and glutathione peroxidase in glioma cells.**

Zhong W, Yan T, Lim R, Oberley LW.

Radiation Research Laboratory, Department of Radiology, The University of Iowa, Iowa City 52242, USA.

Four primary antioxidant enzymes were measured in both human and rat glioma cells. Both manganese-containing superoxide dismutase (MnSOD) and copper-zinc-containing superoxide dismutase (CuZnSOD) activities varied greatly among the different glioma cell lines. MnSOD was generally higher in human glioma cells than in rat glioma cells and relatively higher than in other tumor types. High levels of MnSOD in human glioma cells were due to the high levels of expression of MnSOD mRNA and protein. Heterogeneous expression of MnSOD was present in individual glioma cell lines and may be due to subpopulations or cells at different differentiation stages. Less difference in CuZnSOD, catalase, or glutathione peroxide was found between human and rat glioma cells. The human glioma cell lines showed large differences in sensitivity to the glutathione modulating drugs 1,3-bis (2-chloroethyl)-1-nitrosourea (BCNU) and buthionine sulfoximine (BSO). A good correlation was found between sensitivity to BCNU and the activities of catalase in these cell lines. Only one cell line was sensitive to BSO and this line had low CuZnSOD activity.

PMID: 10641728 [PubMed - indexed for MEDLINE]



**[Expression of human telomerase reverse transcriptase in cervix cancer and its significance]**

[Article in Chinese]

Xi L, Zhu T, Wu P, Xu Q, Huang L, Li KZ, Lu YP, Ma D.

Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China.

**OBJECTIVE:** To investigate the expression of human telomerase reverse transcriptase (hTERT) mRNA and protein in cervix cancer, cervical intraepithelial neoplasia (CIN) and normal cervix. **METHODS:** Expression of hTERT mRNA and the other two subunits of telomerase, human telomerase RNA component (hTR), human telomerase-associated protein (hTP1) was determined by RT-PCR in 3 cervix cancer cell lines, 2 diploid cell lines, 38 cases of cervix cancer, 16 cases of CIN and 20 cases of normal cervix.

Telomerase activity was also examined by telomeric repeat amplification protocol enzyme-linked immunosorbent assay (TRAP-ELISA). Expression of hTERT protein was detected in all the cell lines and 101 cases of paraffinized cervix tissue sections.

**RESULTS:** hTERT mRNA expression was detected in all of the three cervix cancer cell lines, 81.6% of cervix cancer, 37.5% of CIN, 5.0% of normal cervix, while in neither of the two diploid cell lines. The other two subunits of telomerase were prevalently expressed in all of the cell lines and most cervix tissues. There was a strong correlation between hTERT mRNA expression and telomerase activity. Immunostaining also revealed that hTERT protein was expressed in all three cervix cancer cell lines, 65.5% of cervix cancer, 28.0% of CIN and 4.8% of normal cervix. **CONCLUSION:** Up-regulation of hTERT may play an important role in the development of CIN and cervix cancer, hTERT could be used as an early diagnostic biomarker for cervix cancer.

PMID: 16008894 [PubMed - in process]

both general and selected populations. We will have to take the next steps with not only advanced computational tools in hand, but also with a measured

and useful eye towards the clinical context, one that guides us towards the introduction and utilization of genetic information to improve health worldwide.

#### References

- Collins, F.S. (1999) Shattuck lecture – medical and societal consequences of the Human Genome Project. *N. Engl. J. Med.* 341, 28–37
- Risch, N.J. (2000) Searching for genetic determinants in the new millennium. *Nature* 405, 847–856
- Schork, N.J. *et al.* (2000) Single nucleotide polymorphisms and the future of genetic epidemiology. *Clin. Genet.* 58, 250–264
- Collins, F.S. and Guttman, A.E. (2001) Genetics moves into the medical mainstream. *J. Am. Med. Assoc.* 286, 2322–2324
- Brookes, A.J. (2001) Rethinking genetic strategies to study complex diseases. *Trends Mol. Med.* 7, 512–516
- Lupski, J.R. (1998) Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet.* 14, 417–422
- Stankiewicz, P. and Lupski, J.R. (2002) Genome architecture, rearrangements and genomic disorders. *Trends Genet.* 18, 74–82
- Garred, P. *et al.* (1999) Association of mannose-binding lectin gene heterogeneity with severity of lung disease and survival in cystic fibrosis. *J. Clin. Invest.* 104, 431–437
- Chanock, S.J. and Foster, C.B. (1999) SNPing away at innate immunity. *J. Clin. Invest.* 104, 369–370
- Struwing, J.P. *et al.* (1997) The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews. *N. Engl. J. Med.* 336, 1401–1408
- Modan, B. *et al.* (2001) Parity, oral contraceptives, and the risk of ovarian cancer among carriers and noncarriers of a BRCA1 or BRCA2 mutation. *New Engl. J. Med.* 345, 235–240
- Chanock, S.J. (2001) Candidate genes and single nucleotide polymorphisms in the study of human disease. *Disease Markers*, 17, 89–98
- Hill, A.V. *et al.* (1997) Genetic analysis of host-parasite coevolution in human malaria. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 352, 1317–1325
- Hill, A.V. (2001) The genomics and genetics of human infectious disease susceptibility. *Annu Rev Genomics Hum Genet.* 2, 373–400
- Risch, N. and Merikangas, K. (1996) The future of genetic studies of complex human diseases. *Science* 273, 1516–1517
- Dahlman, I. *et al.* (2002) Parameters for reliable results in genetic association studies in common disease. *Nat. Genet.* 30, 149–150
- Pritchard, J.K. (2001) Are rare variants responsible for susceptibility to complex diseases? *Am. J. Hum. Genet.* 69, 124–137
- Foster, C.B. *et al.* (1998) Host defense molecule polymorphisms influence the risk for immune-mediated complications in chronic granulomatous disease. *J. Clin. Invest.* 102, 2146–2155
- Roses, A.D. (2002) Pharmacogenetics place in modern medical science and practice. *Life Sci.* 70, 1471–1480
- Evans, W.E. and Relling, M.V. (1999) Pharmacogenomics: translating functional genomics into rational therapeutics. *Science* 286, 487–491
- Collins, A. *et al.* (1999) Genetic epidemiology of single-nucleotide polymorphisms. *Proc. Natl. Acad. Sci. U. S. A.* 96, 15173–15177
- Reich, D.E. *et al.* (2001) Linkage disequilibrium in the human genome. *Nature* 411, 199–204
- Fallin, D. *et al.* (2001) Genetic analysis of case/control data using estimated haplotype frequencies: application to APOE locus variation and Alzheimer's disease. *Genome Res.* 11, 143–151

# The value of microarray techniques for quantitative gene profiling in molecular diagnostics

Stephen A. Bustin and Sina Dorudi

There has been an explosion of interest in microarray technologies that allow the quantification of whole-genome RNA expression data. The apparent correlation of expression profiles with clinically relevant parameters such as disease outcome has raised expectations with respect to the clinical usefulness of the data generated. Yet the accuracy and biological relevance of these data remain contentious, even in basic research applications. Therefore, numerous issues related to format, quality, validation and interpretation remain to be resolved before microarray profiling can become a diagnostic tool of clinical relevance for routine work.

Published online: 8 May 2002

The current fashion for functional genomics has put the spotlight on microarray technologies that are capable of comprehensive, quantitative RNA

expression analysis, with their promise of a global approach to the quantification of gene expression [1]. However, there are two conflicting opinions on the value of microarrays: they are either promoted as an exploratory-driven replacement for hypothesis-driven biology [2], or they are criticized as expensive fads [3] that cannot substitute for the old-fashioned, low-throughput approach to experimental biology [4].

#### Variability

Unquestionably, microarrays can reveal associations between gene-expression signatures and the biology and outcome of disease, for example by identifying clinically significant subtypes of cancers [5,6]. This has raised expectations that the expression profile of a cell can be used as a diagnostic and/or prognostic aid in cancer management. However, although microarray experiments generate long lists of genes with altered expression, the interpretation of these data depends on the judgement of the investigator performing the experiment [7]. Furthermore, a comparison of the same microarray experiment performed a few weeks apart can demonstrate a considerable lack of reproducibility [8]. This is exacerbated by an apparent lack of robustness of the microarray data, as demonstrated by the variability in the results obtained in different laboratories. Indeed, a comparison of data obtained in independent studies performed with different microarray platforms, for example in lung [9,10] or colorectal [11–13] cancers, shows both similarities and significant differences. Moreover, a comparison of 47 and 98 genes identified from independent studies to be associated with metastasis [14,15] does not reveal a

single common gene. Different methods of data analysis are one cause for the variation in microarray results in the membership of genes on one list but not on another from a similar experiment. For example, although patients with diffuse large B-cell lymphoma could be categorized into groups with differences in overall survival using unsupervised clustering analysis of expression profiles [5], the prognostic power of these gene signatures was not confirmed in a more recent study that used a supervised clustering method of analysis [16]. However, 13 key genes defined in the second study were able to predict disease outcome in both patient series. Good experimental design is crucial, and misleading conclusions can appear in even the most rigorously refereed publications [17].

#### Tissue selection

Normal biopsies contain a range of different cell types, a problem exacerbated in heterogeneous tumour samples that also include normal and inflammatory cells as well as diversely evolved cell populations. Furthermore, normal cells adjacent to a tumour could be phenotypically normal but genotypically abnormal, or could exhibit altered gene expression profiles owing to their proximity to the tumour [18]. This does not present a problem for conventional histopathology, but expression profiling of such biopsies provides a composite of the whole population, and cannot identify expression limited to subpopulations or sections of the biopsy.

The accuracy of quantitative gene profiling is critically dependent on microdissection, especially on the laser capture microdissection (LCM) of individual cells or cell populations [19]. However, because the amount of mRNA becomes a limiting factor, any amplification procedure has the potential to distort the original mRNA distribution pattern [20]. This is important since quantitative expression profiling relies on the preservation of mRNA species distribution in the tissue of interest. New procedures for target amplification are being developed that appear to maintain comparable expression profiles between amplified and non-amplified target [21,22], but results obtained using the different protocols have not been tested for comparability. In addition, is inclusion of labour-intensive and costly steps such as LCM a realistic proposition for routine clinical diagnostics? Even if it is, expression-profile analysis might only refine, rather than revolutionize, current histopathological staging information [23].

#### Compatibility

Fundamental issues of protocol and platform comparability remain to be addressed. There is no standardization of protocols or reagents for sample acquisition, or RNA or cDNA target preparation. This, together with the lack of a universally accepted measure of quality assessment, virtually ensures significant variability with respect to target preparation. The microarray measurement process itself relies on consecutive steps that influence data

reproducibility and comparability: array construction, target labelling, probe-target hybridization, signal detection, normalization and interpretation [24]. The array user generally has no control over quality-control issues affecting fabrication of the DNA microarray platform; however, commercial cDNA microarrays and oligonucleotide chip-based systems have had significant errors in sequence identity [25,26]. The user does have control over the implementation of target labelling, hybridization and detection protocols, but this is likely to result in additional sources of variability, as there are uncertainties associated with different measuring instruments and procedures [27]. Unfortunately, there is no consistency or standardization with respect to any single component to aid comparative analysis.

#### Analysis

The meaningful interpretation of results depends on the appropriateness of the techniques employed for data analysis [28], and the accuracy of microarray profiles requires correct study design, well-characterized statistical goals and careful attention to 'old-fashioned' and proven statistical principles [29]. Since array data typically consist of parallel expression measures for thousands of genes, a combination of analysis of variance (ANOVA) analysis and bootstrap re-sampling of the residuals has been suggested as the most appropriate way of obtaining confidence intervals of the differences in expression levels for each gene [30]. On the other hand, lack of power is an important limitation, as experimental replication at the level of single genes is low and current statistical methods used in microarray analysis have been severely criticized [31]. One can apply stringent criteria to avoid false-positive results, but this is usually at the expense of increasing the false-negative rate [32]. Sophisticated software tools that identify common patterns of expression have been developed and most data are analyzed using clustering techniques. These identify distinct expression patterns and assume functional relationships by grouping genes with similar expression patterns ('guilt by association'), but do not address the relevance of any interrelationship of their expression levels [33]. Analyses also need to take into account the significant specific biases that exist among probes used to interrogate the same gene [34].

#### Normalization

All procedural errors are compounded by continuing uncertainties in the normalization step. Signal intensity is the most critical parameter that influences the reproducibility, specificity and sensitivity of microarray experiments [35], but the low dynamic range of microarrays [36] can bias expression profiles towards low-abundant targets [37]. Furthermore, the expression of housekeeping genes, which are frequently used as internal standards, is variable [38]. Therefore, validation of expression trends in an unrelated, secondary assay remains a critical requirement, and is best accomplished using the real-time reverse transcriptase polymerase chain

Stephen A. Bustin\*  
Sina Dorudi  
Academic Dept of Surgery,  
Barts and the London,  
Queen Mary's School of  
Medicine and Dentistry,  
University of London,  
London, UK E1 1BB.  
\*e-mail: s.a.bustin@  
qmul.ac.uk

Table 1. Microarray databases and analysis tools

Function	Organization	URL
MIAME	MGED Group	<a href="http://www.mged.org/Workgroups/MIAME/miame.html">www.mged.org/Workgroups/MIAME/miame.html</a>
Data exchange protocols: MAGE-ML	MGED Group	<a href="http://www.mged.org/Workgroups/MAGE/mage.html">www.mged.org/Workgroups/MAGE/mage.html</a>
Array-Express database, submission and analysis tools	EMBL-EBI	<a href="http://www.ebi.ac.uk/microarray">www.ebi.ac.uk/microarray</a>
GEO database	NCBI	<a href="http://www.ncbi.nlm.nih.gov/geo">www.ncbi.nlm.nih.gov/geo</a>
ArrayDB database	NHGRI	<a href="http://genome.nhgri.nih.gov/arraydb">http://genome.nhgri.nih.gov/arraydb</a>
GeneX database and tools	NCGR	<a href="http://www.ncgr.org/genex">www.ncgr.org/genex</a>
Protocols and tools	DeRisi lab, University of California (San Francisco)	<a href="http://www.microarrays.org">www.microarrays.org</a>
Software and tools	Biodiscovery	<a href="http://www.biodiscovery.com">www.biodiscovery.com</a>
Software and tools	University of California (Berkeley)	<a href="http://rana.lbl.gov">http://rana.lbl.gov</a>
Reagents	Stratagene	<a href="http://www.stratagene.com">www.stratagene.com</a>

Abbreviations: GEO, gene expression omnibus; EMBL-EBI, European Molecular Biology Laboratory-European Bioinformatics Institute; MAGE-ML, microarray and gene expression markup language; MGED, microarray gene expression data; MIAME, minimum information about a microarray experiment; NCBI, National Center for Biotechnology Information; NCGR, National Center for Genome Resources; NHGRI, National Human Genome Research Institute.

reaction (RT-PCR) with its high dynamic range, bearing in mind the limitations of the technique [39].

#### Relevance

How relevant are the data being obtained? DNA array work is predicated on the assumption that quantification of cellular levels of mRNA transcripts can function as a surrogate for measuring the expression or activity of the protein, which is the final functional product of gene expression. However, the correlation between the number of mRNA and protein molecules is generally not strong enough to predict one value from the measurement of the other [40,41], as there is at least a ten-fold range of protein abundance for mRNAs of a given abundance [42]. Certainly, gene expression levels *per se* tell us very little about the biological function of the gene, its potential clinical impact or its suitability as a drug target. A complete picture can only be obtained by assessing multiple biochemical parameters, as was shown recently in a report that combined microarrays with quantitative proteomics and databases of known physical interactions to build a biochemical pathway, test the hypothesis and refine the model generated [43]. Furthermore, most microarrays are not exhaustive and report on selected subsets of genes. Therefore, it is possible that relevant information will be missed and that, at this early stage, more-exhaustive analyses still need to be carried out using arrays containing as many genes as possible. Third, most arrays are not designed to discriminate between alternatively spliced transcripts of the same gene, or between closely related family members [44]. On the other hand, as several microarray experiments have generated clinically relevant quantitative gene profiles, this question might not be quite as relevant for molecular diagnostics as it is for mechanistic research applications.

#### Minimum information about a microarray experiment

Easy access to DNA sequence databases and the ability to mine sequences gathered from many disparate sources has been a key contributor to the success of the Human Genome Project. mRNA profiling data typically consist of many thousands of measurements

for each array and, unlike DNA sequence data, depend on the context of the individual biological sample and the experimental conditions under which the sample was taken. Hence, access to raw data and the experimental conditions resulting in these data are prerequisites to enable comparative analyses as well as automated data mining [45]. This requires the adoption of standards, although it is worth pointing out that too-early standardization might establish a standard that restricts rather than enables [17]. One proposed solution is adherence to a Minimum Information About a Microarray Experiment (MIAME) standard for array experiment annotation and data representation [46], to provide the minimum information required for unambiguous interpretation and independent verification of microarray data.

#### Repositories

Although it is obvious that the annotated microarray data will have to be stored in accessible locations, it is not surprising in the light of the many unresolved items associated with these data that the usefulness of such repositories has been questioned [47]. Some repositories state that they are MIAME compliant (e.g. Array-Express), whereas others support the spirit of MIAME but are still considering the meaning of MIAME compliance (e.g. GEO). However, anyone used to mining public DNA sequence databases is aware of the numerous inconsistencies and errors associated with the data deposited therein. Hence, it will be crucial to have a readily applicable and transparent means of validating data quality [48].

#### Ontologies and toolsets

Data interpretation represents an enormous challenge for bioinformatics and, to date, there is no consensus on how to merge data from different array types, let alone how to include data from alternative gene expression profiling approaches, such as differential display and serial analysis of gene expression. Hence, there is an urgent need to apply agreed ontologies (i.e. sets of well-defined terms with well-defined relationships) to the description of experimental conditions, samples

and their treatment sources [49]. Also required are integrated toolsets that permit users to download their data into the database, and to use common data-exchange protocols such as Microarray Annotation Markup Language (MAML) to communicate with other databases in order to refine their own array data or for more-general automated data mining. The National Centre for Genome Resources' GeneX™ project aims to create an internet-available relational database of public data derived from multiple technologies, and standardization programmes have been undertaken under the umbrella of the Microarray Gene Expression Data (MGED) group. Table 1 summarizes the main public-repositories analysis tools and a source of reagents for microarray analyses.

## Conclusion

Microarrays are already having a major impact on cancer biology, pharmacology and drug development and, with their rapid development, new technologies, methods of analysis and applications are emerging continuously. However, the major limiting factor in their further application is the current lack of data comparability, which is essential for appropriate comparisons between different array data. Therefore, standards for rigorous data annotation, agreed ontologies and data deposition in public repositories must be agreed if this technology is to be of practical use and become a routine tool for diagnostic and therapeutic applications.

## References

- Duggan, D.J. *et al.* (1999) Expression profiling using cDNA microarrays. *Nat. Genet.* 21, 10–14
- Brown, P.O. and Botstein, D. (1999) Exploring the new world of the genome with DNA microarrays. *Nat. Genet.* 21, 33–37
- Brenner, S. (1999) Sillycon valley fever. *Curr. Biol.* 9, R671
- Modlin, R.L. and Bloom, B.R. (2001) Chip shots – will functional genomics get functional? *Science* 294, 799–801
- Alizadeh, A.A. *et al.* (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403, 503–511
- Pomeroy, S.L. *et al.* (2002) Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature* 415, 436–442
- Mirmics, K. (2001) Microarrays in brain research: the good, the bad and the ugly. *Nat. Rev. Neurosci.* 2, 444–447
- Lander, E.S. (1999) Array of hope. *Nat. Genet.* 21, 3–4
- Garber, M.E. *et al.* (2001) Diversity of gene expression in adenocarcinoma of the lung. *Proc. Natl. Acad. Sci. U. S. A.* 98, 13784–13789
- Bhattacharjee, A. *et al.* (2001) Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc. Natl. Acad. Sci. U. S. A.* 98, 13790–13795
- Alon, U. *et al.* (1999) Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc. Natl. Acad. Sci. U. S. A.* 96, 6745–6750
- Kitahara, O. *et al.* (2001) Alterations of gene expression during colorectal carcinogenesis revealed by cDNA microarrays after laser-capture microdissection of tumor tissues and normal epithelia. *Cancer Res.* 61, 3544–3549
- Takemasa, I. *et al.* (2001) Construction of preferential cDNA microarray specialized for human colorectal carcinoma: molecular sketch of colorectal cancer. *Biochem. Biophys. Res. Commun.* 285, 1244–1249
- Hegde, P. *et al.* (2001) Identification of tumor markers in models of human colorectal cancer using a 19,200-element complementary DNA microarray. *Cancer Res.* 61, 7792–7797
- Yanagawa, R. *et al.* (2001) Genome-wide screening of genes showing altered expression in liver metastases of human colorectal cancers by cDNA microarray. *Neoplasia* 3, 395–401
- Shipp, M.A. *et al.* (2002) Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat. Med.* 8, 68–74
- Foster, W.R. and Huber, R.M. (2002) Current themes in microarray experimental design and analysis. *Drug Discov. Today* 7, 290–292
- Deng, G. *et al.* (1996) Loss of heterozygosity in normal tissue adjacent to breast carcinomas. *Science* 274, 2057–2059
- Bowtell, D.D. (1999) Options available – from start to finish – for obtaining expression data by microarray. *Nat. Genet.* 21, 25–32
- Van Gelder, R.N. *et al.* (1990) Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc. Natl. Acad. Sci. U. S. A.* 87, 1663–1667
- Wang, E. *et al.* (2000) High-fidelity mRNA amplification for gene profiling. *Nat. Biotechnol.* 18, 457–459
- Vernon, S.D. *et al.* (2000) Reproducibility of alternative probe synthesis approaches for gene expression profiling with arrays. *J. Mol. Diagn.* 2, 124–127
- Lakhani, S.R. and Ashworth, A. (2001) Microarray and histopathological analysis of tumours: the future and the past? *Nat. Rev. Cancer* 1, 151–157
- Schuchhardt, J. *et al.* (2000) Normalization strategies for cDNA microarrays. *Nucleic Acids Res.* 28, E47
- Halgren, R.C. *et al.* (2001) Assessment of clone identity and sequence fidelity for 1189 IMAGE cDNA clones. *Nucleic Acids Res.* 29, 582–588
- Knight, J. (2001) When the chips are down. *Nature* 410, 860–861
- Brown, C.S. *et al.* (2001) Image metrics in the statistical analysis of DNA microarray data. *Proc. Natl. Acad. Sci. U. S. A.* 98, 8944–8949
- Quackenbush, J. (2001) Computational analysis of microarray data. *Nat. Rev. Genet.* 2, 418–427
- Kerr, M.K. and Churchill, G.A. (2001) Statistical design and the analysis of gene expression microarray data. *Genet. Res.* 77, 123–128
- Kerr, M.K. *et al.* (2000) Analysis of variance for gene expression microarray data. *J. Comput. Biol.* 7, 819–837
- Szabo, A. *et al.* (2002) Variable selection and pattern recognition with gene expression data generated by the microarray technology. *Math. Biosci.* 176, 71–98
- Geschwind, D.H. (2001) Sharing gene expression data: an array of options. *Nat. Rev. Neurosci.* 2, 435–438
- Woolf, P.J. and Wang, Y. (2000) A fuzzy logic approach to analyzing gene expression data. *Physiol. Genomics* 3, 9–15
- Li, C. and Wong, W.H. (2001) Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc. Natl. Acad. Sci. U. S. A.* 98, 31–36
- Yang, M.C. *et al.* (2001) A statistical method for flagging weak spots improves normalization and ratio estimates in microarrays. *Physiol. Genomics* 7, 45–53
- Mills, J.C. and Gordon, J.I. (2001) A new approach for filtering noise from high-density oligonucleotide microarray datasets. *Nucleic Acids Res.* 29, E72
- Baldi, P. and Long, A.D. (2001) A Bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes. *Bioinformatics* 17, 509–519
- Eickhoff, B. *et al.* (1999) Normalization of array hybridization experiments in differential gene expression analysis. *Nucleic Acids Res.* 27, E33
- Bustin, S.A. (2002) Quantification of mRNA using real-time RT-PCR: trends and problems. *J. Mol. Endocrinol.* 28 (in press)
- Anderson, L. and Seilhamer, J. (1997) A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* 18, 533–537
- Cygi, S.P. *et al.* (1999) Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* 19, 1720–1730
- Futcher, B. *et al.* (1999) A sampling of the yeast proteome. *Mol. Cell. Biol.* 19, 7357–7368
- Ideker, T. *et al.* (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 292, 929–934
- Freeman, W.M. *et al.* (2000) Fundamentals of DNA hybridization arrays for gene expression analysis. *Biotechniques* 29, 1042–1055
- Hayes, A. (2000) The second international meeting on Microarray Data Standards, Annotations, Ontologies and Databases. *Yeast* 17, 238–240
- Brazma, A. *et al.* (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat. Genet.* 29, 365–371
- Hess, K.R. *et al.* (2001) Microarrays: handling the deluge of data and extracting reliable information. *Trends Biotechnol.* 19, 463–468
- Becker, K.G. (2001) The sharing of cDNA microarray data. *Nat. Rev. Neurosci.* 2, 438–440
- Ashburner, M. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25–29

# 17

## Proteomics

**Kathryn Lilley, Azam Razzaq and Michael J. Deery**

### 17.1 Introduction

There are now numerous organisms whose genome sequences are known, for example *Arabidopsis* (*Arabidopsis* Genome Initiative, 2000), *Drosophila* (Adams *et al.*, 2000), human (International Human Genome Sequencing Consortium, 2001) and rice (Yu *et al.*, 2002). The prediction of open-reading frames from these genomic sequences has enabled the comprehensive identification of many putative protein sequences. These proteins can be arranged into three categories, namely those of known function, those with recognisable motifs and hence a vague idea of function, and those with no sequence similarities to any protein (Gabor Miklos and Malenszka, 2001). Many proteins reside in this latter 'functional vacuum' which could represent as much as 30% of the predicted proteins. Determining protein function is key to understanding cellular mechanism. Studying how protein expression is modulated in response to a given set of circumstances, such as infection, disease, developmental stage, senescence or response to drugs, will facilitate the elucidation of disease pathways and thus provide a mechanism for diagnosis and therapy.

DNA chips (mRNA profiling studies) can contribute to the study of gene expression in response to a particular biological perturbation. However, the extrapolation that changes in transcript level will also result in corresponding changes in protein amount or activity cannot always be made. To understand fully, we need integrated data sets from a variety of protein expression studies, providing information on relative abundances, sub-cellular locations, protein complex formation and the profiling of isoforms generated by either alternate mRNA splicing or post-translational modifications. Proteomics is the word now

commonly used to describe the discipline associated with the acquisition of these data sets.

## 17.2 Definitions and applications

Before embarking on an overview of the techniques used in the field of proteomics, it is necessary to define a few terminologies; proteomics, functional genomics, structural genomics and post-genomics are terms that have crept into the scientific vocabulary with alarming stealth and are used freely, and in some cases interchangeably, especially in cases where the use of such 'buzz words' is likely to increase the attendance at departmental seminars.

- **Genome:** A genome represents the entire DNA content in a particular cell, whether or not it is coding, non-coding, or is located either chromosomally or extra-chromosomally.
- **Genomics:** Genomics is the study of the genome, interrogating the complete genome sequence using both DNA and RNA methodologies.
- **Proteome:** The proteome represents the complete set of proteins encoded by the genome.
- **Proteomics:** Proteomics is the study of the proteome and investigates the cellular levels of all the isoforms and post-translational modifications of proteins that are encoded by the genome of that cell under a given set of circumstances.
- **Functional genomics:** This is the study of the functions of genes and their inter-relationships.

Whilst a genome is more or less static, the protein levels in a particular cell can change dramatically as genes get turned on and off during the cell's response to its environment. The proteome originally was defined over seven years ago as 'all the proteins coded by the genome of an organism' (Wasinger *et al.*, 1995). Nowadays the term 'proteomics' is used more widely and implies an effort to link structure to function by whatever means are appropriate. We can expect the definition to change again with time as the field and the investigator's view of it evolves.

## 17.3 Stages in proteome analysis

Proteomic analysis can ascertain function either by looking for changes in the expression of either all or a subset of proteins, or by identifying binding partners for particular proteins and seeing how their interaction is affected by biological perturbation. Whatever the rationale of the investigation, or the number of proteins involved, the study of the proteome can be broken down into the following stages of analysis.

# Importance of replication in microarray gene expression studies: Statistical methods and evidence from repetitive cDNA hybridizations

Mel-Ling Ting Lee<sup>\*†‡§</sup>, Frank C. Kuo<sup>†¶</sup>, G. A. Whitmore<sup>||</sup>, and Jeffrey Sklar<sup>†¶</sup>

<sup>\*</sup>Departments of Medicine and <sup>†</sup>Pathology, Brigham and Women's Hospital, Boston, MA 02115; <sup>‡</sup>Harvard Medical School, Boston, MA 02115; <sup>§</sup>Biostatistics Department, Harvard School of Public Health, Boston, MA 02115; and <sup>¶</sup>Faculty of Management, McGill University, Montreal, Quebec, Canada H3A 1G5

Edited by Bradley Efron, Stanford University, Stanford, CA, and approved June 23, 2000 (received for review March 13, 2000)

We present statistical methods for analyzing replicated cDNA microarray expression data and report the results of a controlled experiment. The study was conducted to investigate inherent variability in gene expression data and the extent to which replication in an experiment produces more consistent and reliable findings. We introduce a statistical model to describe the probability that mRNA is contained in the target sample tissue, converted to probe, and ultimately detected on the slide. We also introduce a method to analyze the combined data from all replicates. Of the 288 genes considered in this controlled experiment, 32 would be expected to produce strong hybridization signals because of the known presence of repetitive sequences within them. Results based on individual replicates, however, show that there are 55, 36, and 58 highly expressed genes in replicates 1, 2, and 3, respectively. On the other hand, an analysis by using the combined data from all 3 replicates reveals that only 2 of the 288 genes are incorrectly classified as expressed. Our experiment shows that any single microarray output is subject to substantial variability. By pooling data from replicates, we can provide a more reliable analysis of gene expression data. Therefore, we conclude that designing experiments with replications will greatly reduce misclassification rates. We recommend that at least three replicates be used in designing experiments by using cDNA microarrays, particularly when gene expression data from single specimens are being analyzed.

Although the high-throughput technology now available enables genetic researchers to study expression for thousands of genes simultaneously, experiments by using microarrays may be costly and time consuming. The manufacturers of microarray equipment do not stress the need for replication of studies. Production of arrays can be slow and the supply limited. As a result, most current molecular genetic studies that use microarray technology are sometimes done without replication. However, statistical analyses in many settings have demonstrated that important insights into the nature of inherent variability are obtained by the replication of experiments.

In Section 1, we report the design of a controlled experiment involving replication of cDNA hybridizations. The study was conducted to investigate inherent variability in gene expression data and the extent to which replication in an experiment produces more consistent and reliable findings. In Sections 2.1 and 2.2, we introduce statistical models to describe the probability that an mRNA is contained in the target sample tissue, converted to probe, and ultimately detected on the slide as an observed expression. We use a mixed normal distribution to model the distribution of observed gene expressions. In Sections 2.3 and 2.4, we conduct a separate analysis for each replicate. In Sections 2.5 and 2.6, we introduce a model to provide a joint analysis based on the combined data collected from all replicates. In Section 2.7, we consider the reliability of the classification of gene expression as a function of the number of replicates.

Our results show that any single microarray output is subject to substantial variability. By pooling data from replicates, we can provide a more reliable classification of gene expression. Therefore, we conclude that designing experiments with replications will greatly reduce misclassification rates. We recommend that at least three replicates be used in designing experiments using cDNA microarrays. Although our results depend on specific instruments and techniques, the statistical models and methods that we propose in this article can be applied in general settings.

## 1. Materials and Methods

In this section, we provide a brief description of our experimental process. To check the consistency of microarray experiments, we conducted a study to investigate whether the unevenness of the surfaces of glass slides, the locations of cDNA spots on the slides, and other aspects of a microarray experiment may produce variation in measurements of transcriptions. To test these variables of cDNA microarrays generated in our facility, we printed triplicates of 288 cDNA sets (288 elements per set) at 3 locations on the same slide and performed hybridization experiments with probes from 1 source. By comparing the signals from these triplicates, we hoped to learn about the reproducibility of the array process and whether seemingly minor factors, such as the location of the spots in the array, can affect the outcome of analyses. Of the 288 genes considered in this experiment, 32 would be expected *a priori* to appear highly expressed because of structural features within the genes, namely Alu repeats that should crosshybridize to similar sequences widely distributed among expressed and nonexpressed portions of the genome.

**1.1. Generation of Array-Ready cDNAs.** Frozen glycerol stocks of *Escherichia coli* containing individual cDNA clones in the IMAGE consortium distributed in 384-well plates were purchased from Genome Systems, St. Louis. Individual bacterial clones were selected and distributed into 96-well plates. Amplifications of DNA by PCR with primers specific to the vector sequences flanking the insert cDNA were performed in 96-well PCR plates in a Perkin-Elmer 9600 thermocycler in 50- $\mu$ l reactions containing  $\times 1$  PCR buffer (Promega), 1.5 mM MgCl<sub>2</sub>, 0.2 mM dNTPs, 10 pmol of each primer, 5 units of *Taq*

This paper was submitted directly (Track II) to the PNAS office.

<sup>§</sup>To whom reprint requests should be addressed at: Channing Laboratory, BWH/HMS, 181 Longwood Avenue, Boston, MA, 02115-5804. E-mail: stmel@channing.harvard.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.



polymerase, and 0.5  $\mu$ l of the bacterial culture. The annealing was at 55°C for 20 sec, and the extension was at 72°C for 90 sec for 30 cycles. Five microliters of the PCR reactions was analyzed in a 1% agarose gel to verify the success of PCR. The DNA in the remaining 45  $\mu$ l was precipitated with addition of 45  $\mu$ l of isopropanol and resuspended in 15  $\mu$ l of  $\times 3$  SSC. Note that 32 of the 288 genes contained Alu repetitive sequences and therefore were expected to show a high level of signal because of crosshybridization of Alu containing messages.

**1.2. Printing of the cDNAs on Glass Slides.** The array-ready cDNAs in 96-well plates were loaded into a 417 arrayer manufactured by Genetic Microsystems (Woburn, MA). Poly-L-lysine-coated slides from Sigma were used as the solid support for construction of the array. Triplicate arrays were placed on the slides at 6-mm intervals. After printing, the slides were hydrated over a steam bath and the DNA UV-crosslinked onto the slides. After blocking the slides with succinic anhydride, the DNA was denatured by boiling for 2 min, and the slides were dehydrated with ethanol.

**1.3. Preparation of Fluorescently Labeled Probes and Hybridization to Glass Slides.** Total RNA was isolated with Trizol reagents (Life Technologies, Grand Island, NY) from human tissue specimens obtained during surgical procedure. Fifty micrograms of total RNA was annealed to oligo(dT) and reverse transcribed in the presence of Cy3-labeled dUTP. The resulting cDNA was precipitated with ethanol, resuspended in 20  $\mu$ l of hybridization solution [50% formamide/ $\times 5$  SSC/0.5% SDS/1  $\mu$ g of CoT1 DNA/10  $\mu$ g of yeast tRNA/10 pmol of poly(dA)], heat denatured, applied to the slide, and sealed under a coverslip. The slide was placed in a humidified chamber at 42°C overnight. The washing was in  $\times 0.5$  SSC/0.2% SDS at room temperature twice for 5 min each, followed by three washes with  $\times 0.2$  SSC at room temperature for 2 min each. The slide was dried and scanned with a 418 array reader from Genetic Microsystems. The resulting image was quantified by using the software program SCANALYZE (1). The fluorescence of the Cy3 label is carried on Channel 1. Cy5 was not used in this experiment, and hence Channel 2 carried only background noise.

## 2. Statistical Model and Analytical Approach

For gene  $g$  in experimental replicate  $j$ , where  $g = 1, \dots, G$ ,  $j = 1, \dots, J$ , let  $X_{kj}$  denote the median of the set of background-corrected single pixel values of Channel 1 to Channel 2 fluorescence for all pixels within the fluorescence spot. This measure is denoted by MRAT in Eisen *et al.* (2) and Eisen (1). We take the natural logarithm of MRAT as  $Y_{kj} = \ln(X_{kj})$  and refer to  $Y_{kj}$  as a log-ratio. In this experiment, three replications of expression measurements for 288 gene probes were obtained under the same experimental conditions from the same human tissue sample. Thus,  $G = 288$  and  $J = 3$ .

**2.1. The Probability of Observing Expressed Genes.** Consider any one replicate  $j$  among the three experimental replicates  $j = 1, 2, 3$ . Let  $\mathcal{E}_g$  represent the event that mRNA for gene  $g$  in the array is contained in the target sample tissue. In advance of observing the gene expression data, we attach a prior probability  $Pr(\mathcal{E}_g) = p$  to this event for each gene  $g$  that is under consideration. The fact that  $p$  is not indexed by  $g$  implies that, in advance of considering the experimental data, we are uniformly ignorant about whether any particular gene is contained in the sample tissue. We denote the complement of event  $\mathcal{E}_g$  by  $\bar{\mathcal{E}}_g$ .

For a gene to be detected on the slide, three hurdles must be cleared. First, the mRNA must be part of the sample from which the probe is prepared. Second, some of the mRNA in the sample must be converted to probe. Third, some of the probe must be detected by the cDNAs deposited on the slide. If any one of these

Table 1. Separate analysis for each experimental replicate

Parameter	Replicate		
	$j = 1$	$j = 2$	$j = 3$
$p$	0.285	0.124	0.274
$\mu_{Uj}$	0.364	0.410	0.442
$\mu_{Ej}$	0.968	2.203	1.233
$\sigma_{Uj}^2$	0.070	0.076	0.062
$\sigma_{Ej}^2$	1.186	0.114	1.079

Parameter estimates of the mixed normal model (Eq. 1).

hurdles is not cleared, the gene cannot be expressed in the microarray data.

The log-ratio  $Y_{kj}$  for gene  $g$  in replicate  $j$  will have two distinct distributions, depending on whether gene  $g$  is contained in the sample tissue. First, if mRNA from gene  $g$  is not in the sample tissue (i.e., event  $\bar{\mathcal{E}}_g$ ), its measured expression should reflect only experimental error. In this case, we assume that  $Y_{kj}$  is normally distributed as  $N(\mu_{Uj}, \sigma_{Uj}^2)$ , where subscript  $U_j$  refers to the anticipated outcome of being *unexpressed*. We denote the corresponding probability density function of conditional variable  $Y_{kj}|\bar{\mathcal{E}}_g$  by  $f_{Uj}(y)$ . Observe that the distribution parameter values may vary with the replicate  $j$  that is under consideration. On the other hand, if gene  $g$  is in the sample tissue (i.e.,  $\mathcal{E}_g$ ) and should therefore be detected on the slide, we assume that  $Y_{kj}$  is distributed as  $N(\mu_{Ej}, \sigma_{Ej}^2)$ , where subscript  $E_j$  refers to the anticipated outcome of being *expressed*. We denote the corresponding probability density function of the conditional variable  $Y_{kj}|\mathcal{E}_g$  by  $f_{Ej}(y)$ . Again, we note that the parameters may vary with the replicate  $j$ . By definition, we require  $\mu_{Uj} < \mu_{Ej}$ . For event  $\bar{\mathcal{E}}_g$ ,  $Y_{kj}$  is a measurement reflecting only background noise, or inherent experimental error. For event  $\mathcal{E}_g$ , measurement  $Y_{kj}$  reflects the actual expression of gene  $g$  in the sample tissue, obscured to some degree by the presence of background noise.

**2.2. A Mixture Model for the Distribution of Observed Log Ratios.** Given the complementary events  $\mathcal{E}_g$  and  $\bar{\mathcal{E}}_g$  for any gene  $g$ , the observed log-ratio  $Y_{kj}$  for replicate  $j$  will be distributed according to the following mixed normal probability density function.

$$f_j(y) = pf_{Ej}(y) + (1 - p)f_{Uj}(y). \quad [1]$$

A simple manipulation of the two components of Eq. 1 gives posterior probabilities for whether gene  $g$  is expressed in the sample tissue based on a reading  $Y_{kj} = y$  in replicate  $j$ . Specifically, if the microarray reading for the log-ratio of gene  $g$  is  $Y_{kj} = y$  in replicate  $j$ , the posterior probability that the reading reflects expression of gene  $g$  in the sample tissue (and not simply background noise) is given by

$$Pr\{\mathcal{E}_g | Y_{kj} = y\} = \frac{pf_{Ej}(y)}{f_j(y)}. \quad [2]$$

**2.3. Separate Analysis for Each Replicate.** We now examine the problems of estimating the parameters  $p$ ,  $\mu_{Uj}$ ,  $\sigma_{Uj}^2$ ,  $\mu_{Ej}$ , and  $\sigma_{Ej}^2$  for model 1, interpreting the parameter estimates and using them to estimate the posterior probabilities in Eq. 2.

First we solved for the maximum likelihood estimates of the unknown parameters based on model 1. The estimates were calculated separately for the three replications to see how stable the results are from one replicate to another. The parameter estimates appear in Table 1.

The estimates for replicate 2 in Table 1 are sharply different from those for the other two replicates. The estimate of mean parameter  $\mu_{Ej}$  is much larger than for replicates 1 and 3, and the estimates of variance parameter  $\sigma_{Ej}^2$  and probability  $p$  are much

Table 2. Posterior probability of expression in sample tissue

Gene $g$	Replicate 1		Replicate 2		Replicate 3	
	$Y_{g1} = y$	$Pr(\mathcal{E}_g   Y_{g1} = y)$	$Y_{g2} = y$	$Pr(\mathcal{E}_g   Y_{g2} = y)$	$Y_{g3} = y$	$Pr(\mathcal{E}_g   Y_{g3} = y)$
1	2.043	1.0000	1.6934	0.5000	2.6251	1.0000
2	0.0043	0.1336	0.5551	0.0000	0.6874	0.1134
3	0.4940	0.0877	0.3791	0.0000	0.5065	0.0682
17	0.6646	0.1404	0.2662	0.0000	1.7204	1.0000
18	2.4397	1.0000	2.3081	1.0000	2.2481	1.0000
19	2.2331	1.0000	2.0549	1.0000	2.5257	1.0000

Log ratios  $Y_{gj} = y$  and estimates of posterior probabilities  $Pr(\mathcal{E}_g | Y_{gj} = y)$  for a few illustrative genes  $g$ , for replicates  $j = 1, 2, 3$ .

smaller. It is unclear why replicate 2 is so different from the others, but it serves to remind us that replication does not ensure duplication of results, a fact that cannot be quantified when replication is not used. We also note in Table 1 that the estimate of  $p$  varies greatly from one replicate to another. Recall in our controlled experiment that only 32 of the 288 genes (fraction 0.111) should be classified as expressed. Thus, the estimates of  $p$  provided by replicates 1 and 3 are much too large.

We turn next to estimates of the posterior probabilities (Eq. 2). Table 2 summarizes a representative fragment of the results. We see generally that the posterior probability clearly indicates whether a gene is expressed in the sample tissue and that the results are quite uniform across the three replications. There are occasions, however, as illustrated by the results for gene no. 17, where the three replications do not give uniform results. Replicate 3 for this gene gives a very large posterior probability (1.0000) to the expression event  $\mathcal{E}_g | Y_{g3} = y$ , whereas the other two replicates give smaller probabilities (0.1404 and 0.0000).

**2.4. Checking the Consistency of Results from the Three Replicates.** We next study the extent to which the three replications, analyzed separately, provide consistent classification with respect to gene expression. Using the posterior expression probabilities (such as those in Table 2) for each replicate  $j$ , we will classify a gene  $g$  as being expressed if  $Pr(\mathcal{E}_g | Y_{gj} = y)$  is larger than 0.5 and as not being expressed otherwise. This classification is done independently for each replicate.

Table 3 contains the results of this classification process. Table 3 *Left* shows a three-way crossclassification, whereas, for ease of interpretation, Table 3 *Right* shows the three two-way crossclassifications corresponding to the three pairs of replicates. If the replicates were perfectly consistent, only two cells of Table 3 *Left* would have counts, namely, the cell counting unexpressed genes in all three replicates and the cell counting expressed genes in all three replicates. In fact, however, all of the cells in the table have counts, and four of these are sizeable. This is evidence that the replicates are not perfectly consistent. As one illustration of inconsistency, we note in Table 3 *Left* that 23 genes classified as

expressed in replication 3 are classified as unexpressed in replications 1 and 2. As another illustration, we note in Table 3 *Right* that the numbers of genes classified as expressed in the three replicates are 55, 36, and 58, respectively. As 32 of the 288 genes should be classified as expressed, we are again reminded by these results that replicates 1 and 3 are providing a large number of false positives.

To model the count data in Table 3, we again postulate a prior probability  $p$  that any given gene is expressed in the sample tissue. As discussed earlier, mRNA in the tissue must clear two further hurdles to appear "expressed" on the microarray slide. It must be converted to probe and hybridized to the cDNAs that are deposited on the slide. The stochastic behavior of this mechanical process can cause replications to differ. We let  $r_j$  denote the conditional probability that a gene will be classified as "expressed" in replicate  $j$ , where  $j = 1, \dots, J$ , and assume that the corresponding conditional events are independent from one replicate to another. We also postulate that, by chance, a gene that is not expressed in the sample tissue may appear expressed on the slide because of background noise or other experimental artifacts. The conditional probability of such a spurious indication will be denoted by  $s_j$  for the  $j$ th replicate,  $j = 1, \dots, J$ . Again, we assume that these conditional events are independent among the replicates. We can now estimate these probabilities from the count data in Table 3 using the method of maximum likelihood applied to the following likelihood function

$$\prod_g f(w_{g1}, \dots, w_{gJ}) = \prod_g \{pP_g + (1-p)Q_g\}, \quad [3]$$

where

$$P_g = \prod_{j=1}^J [r_j^{w_{gj}}(1-r_j)^{1-w_{gj}}],$$

$$Q_g = \prod_{j=1}^J [s_j^{w_{gj}}(1-s_j)^{1-w_{gj}}],$$

Table 3. Comparing results of a separate analysis for each replicate

Three-way crossclassification						Three two-way crossclassifications											
Replicate 1	Replicate 3																
	U		E														
	Replicate 2		Replicate 2														
	U	E	U	E	Total	Replicate 1	U	E	Total	Replicate 1	U	E	Total	Replicate 2	U	E	Total
U	207	2	23	1	233	U	230	3	233	U	209	24	233	U	226	26	252
E	19	2	3	31	55	E	22	33	55	E	21	34	55	E	4	32	36
Total	226	4	26	32	288	Total	252	36	288	Total	230	58	288	Total	230	58	288

Crossclassification of the log-ratio for three replicates analyzed separately. A gene is classified as expressed if its posterior probability  $Pr(\mathcal{E}_g | Y_{gj} = y)$  exceeds 0.5 and as unexpressed otherwise.  
U, unexpressed; E, expressed.

and  $w_{gj}$  denotes the indicator variable for whether gene  $g$  is classified as expressed in replicate  $j = 1, \dots, J$ . For our experiment, the number of replicates is  $J = 3$ , and the maximum likelihood estimates based on the data in Table 3 are  $\hat{p} = 0.117$ ,  $\hat{p}_1 = 1.000$ ,  $\hat{p}_2 = 0.974$ ,  $\hat{p}_3 = 0.946$ ,  $\hat{s}_1 = 0.084$ ,  $\hat{s}_2 = 0.013$ , and  $\hat{s}_3 = 0.103$ . We note that  $\hat{p}_1$  was numerically so close to the value 1 that it was set to 1 for the computation.

The probability estimates reveal several points of interest for microarray studies. First, the experimental design purposely selected 32 of 288 genes to be expressed, which is the exact fraction  $p = 0.111$ . Hence, the statistical analysis has reliably reproduced this fraction in the estimate of  $p$ . Second, the estimates of  $r_j$  show that (i) it is not a certainty that an expressed gene will be classified as "expressed" on the slide, and (ii) the probability of doing so can vary from one experimental execution to another. Third, the estimates of  $s_j$  show that "ghost" indications of genes (i.e., false positives) can occur with a frequency as large as 10% in a single experiment.

**2.5. A Model for Analyzing the Combined Data from All Replicates.** We now seek to describe the microarray data from the three replications by a single model that will support a joint analysis. We use the following two-way linear model as a general statistical model for log-ratio data.

$$Y_{gj} = \mu + \alpha_g + \beta_j + \gamma_{gj} + e_{gj}, \text{ for } g = 1, \dots, G, j = 1, \dots, J. \quad [4]$$

Here  $E(Y_{gj}) = \mu + \alpha_g + \beta_j + \gamma_{gj}$  is the mean log-ratio for gene  $g$  under experimental condition  $j$ . The component  $\mu$  is the overall mean log-ratio for all genes and experimental conditions,  $\alpha_g$  is the main effect for gene  $g$ ,  $\beta_j$  is the main effect for experimental condition  $j$ , and  $\gamma_{gj}$  is an interaction term that reflects differential gene expression for gene  $g$  under experimental condition  $j$ . In this particular context, the experimental condition  $j$  refers to replicate  $j$ . The term  $e_{gj}$  is a random error which, by definition, has a mean of zero. We assume that the error terms are independent, but we have no need in this study to make any assumption about their distributional form.

Following our earlier assumption for individual replicates, we define the main effect  $\alpha_g$  for gene  $g$  in model 4 as a random effect that follows one of two distinct distributions according to whether gene  $g$  is expressed in the sample tissue (event  $\mathcal{E}_g$ ). The distribution of  $\alpha_g$  therefore follows a mixed normal model

$$f(a) = pf_E(a) + (1 - p)f_U(a) \quad [5]$$

where we now use symbol  $a$  in place of  $y$  for the variable notation in model 1. We assume that the  $\alpha_g$  are independent effects for different genes  $g$ . We start our study by estimating the overall mean, the main effects for genes, and the main effects for experimental replicates in model 4, as follows.

$$\hat{\mu} = \bar{Y}_{..} \quad [6a]$$

$$\hat{\alpha}_g = \bar{Y}_{.g} - \bar{Y}_{..} \quad [6b]$$

$$\hat{\beta}_j = \bar{Y}_{.j} - \bar{Y}_{..} \quad [6c]$$

where  $\bar{Y}_{..}$ ,  $\bar{Y}_{.g}$ , and  $\bar{Y}_{.j}$  denote average log-ratios for all  $j$ , all  $g$ , and all pairs  $(g, j)$ , respectively.

It is conceivable that all of the effects in model 4 are random. The estimates in Eq. 6, however, are standard fixed-effect estimates. We choose these estimators because they are inherently free of any distributional assumption. In particular, the estimates  $\hat{\alpha}_g$  provided by Eq. 6b are fixed-effect estimates that do not depend on the assumption of a normal mixture distribution. We now use these  $\hat{\alpha}_g$  to estimate the parameters of the mixture distribution in Eq. 5 and subsequently use them again to examine

Table 4. Analysis of the combined data from all three replicates

Parameter	Estimate	Est. Std. Err.
$p$	0.118	0.013
$\mu_U$	-0.204	0.039
$\mu_E$	1.524	0.058
$\sigma_U^2$	0.044	0.003
$\sigma_E^2$	0.126	0.036

Parameter estimates of the mixed normal model (Eq. 5) derived from the estimated main effects for genes  $\hat{\alpha}_g$ .

the assumption of a normal mixture distribution. The parameter estimates of the mixed normal model 5 appear in Table 4, together with estimated standard errors. The standard errors are calculated from 100 bootstrap samples.

**2.6. Analysis Results for the Combined Data.** As the main effects for genes are now estimated from three replications, the results are more sharply delineated than they are in Table 1, where the parameter estimates are calculated separately for each replicate. First, we see that the estimate of  $p = 0.118$  is very close to the known proportion of expressed genes in the sample tissue (32 of 288) and almost identical to the corresponding estimate derived from the count data in Table 3. Second, the estimates of the mean parameters  $\mu_U$  and  $\mu_E$  are well separated. Third, the variance estimates  $\sigma_U^2$  and  $\sigma_E^2$  are smaller than those obtained in separate analyses as listed in Table 1. In fact, they would be expected to be smaller by a factor of about 3. Fourth, the estimate of variance parameter  $\sigma_U^2$  is smaller than that of  $\sigma_E^2$ . This difference is expected by the fact that, in the event of no expression (i.e., event  $\bar{\mathcal{E}}_g$ ), variance parameter  $\sigma_U^2$  reflects the variability of the log-ratio of background noise on two channels. In the event of gene expression (i.e., event  $\mathcal{E}_g$ ), variance parameter  $\sigma_E^2$  reflects two sources of variability: (i) the log-ratio of background noise on two channels, and (ii) the logarithm of gene expression itself.

The posterior probability that gene  $g$  is expressed, given the value of  $\hat{\alpha}_g$ , can be calculated for each gene by using Eq. 2 with  $Y_g$  replaced by  $\hat{\alpha}_g$ . These posterior probabilities are all either close to 1 or close to zero. In fact, classifying the genes according to whether this probability is greater than 0.5, it is found that only 2 of the 288 genes are incorrectly classified as to whether they are expressed. Hence, based on the combined data, the classification gives only two false positives and no false negatives. Specifically, genes nos. 75 and 185 are classified as expressed when they were not included in the experimental set of genes. In contrast, recall from Table 3 that the individual replicates were far from perfect in their ability to classify genes.

Fig. 1 *a* and *b* show normal probability plots of the  $\hat{\alpha}_g$  for the genes classified as expressed and unexpressed, respectively. According to the mixed normal model, these two plots should both be normal if the classification were perfect. The evidence seems quite supportive of the normality assumption in both plots. For the genes classified as expressed, there is some evidence of values being clustered. For those classified as unexpressed, there may be a little contamination of the normal distribution, leading to a few outlying observations relative to a pure normal distribution. Fig. 2 shows an overlay of a histogram of the  $\hat{\alpha}_g$  and the mixed normal probability density function, as described in Eq. 5, based on the parameter estimates in Table 4. A comparison of the histogram and the density function shows that the mixed normal model is quite reasonable. Both the histogram and density function show that the expressed and unexpressed genes occupy well-separated locations on the scale. Note, however, the sharp difference in variability of the two component distributions.

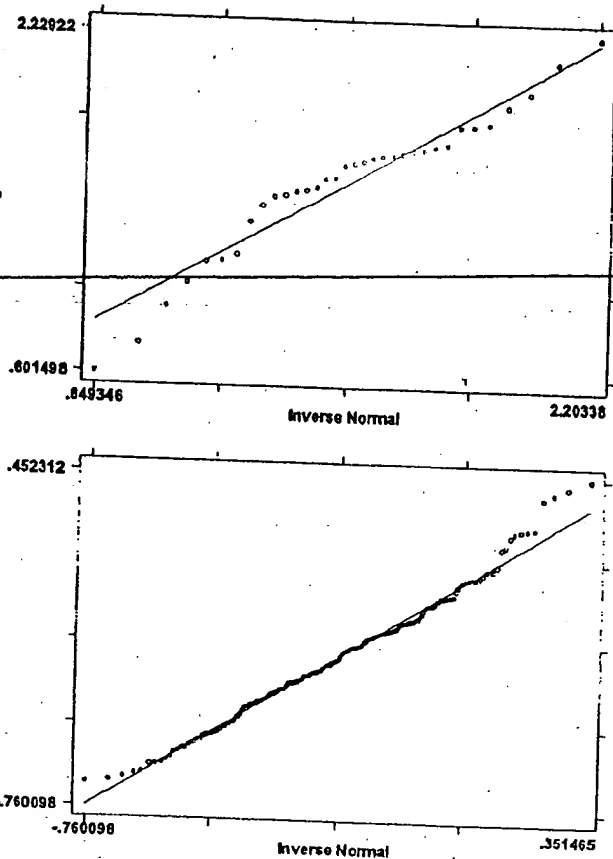


Fig. 1. (a), Normal probability plot of main effect estimates for expressed genes. (b), Normal probability plot of main effect estimates for unexpressed genes.

The interaction terms  $\gamma_{ij}$  in model 4 reflect differential gene expression among the experimental conditions and can be estimated as fixed effects, as follows.

$$\hat{\gamma}_{ij} = Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..} \quad [7]$$

As the experimental conditions here represent replicates, the estimates in Eq. 7 should reflect simply the random noise contributed by the error terms  $e_{ij}$ . We have discovered, however, that the replicates are not true duplicates and that some genes

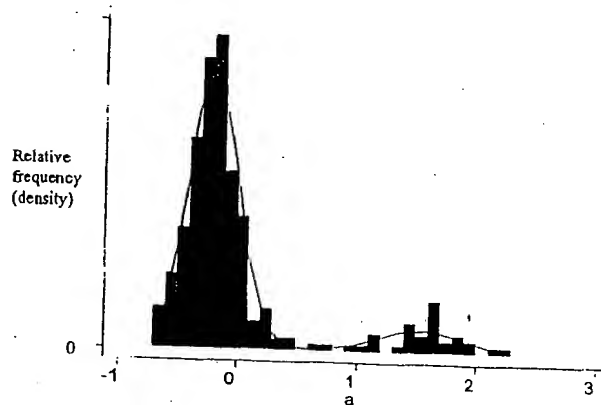


Fig. 2. Overlay of a histogram and mixed normal p.d.f. for gene expression main effect.

may be classified as expressed in one or two replicates but not in all three. The estimates of  $\hat{\gamma}_{ij}$  in these cases therefore indicate a differential expression of the genes. We do not need to study these estimates further because Table 3 describes the patterns of inconsistent expression among the three replicates. We note, however, that in microarray investigations of multiple tissues (or other varying experimental conditions), the estimates of differential expression in Eq. 7 are of central scientific interest in determining which genes are truly present in some tissues but not in others.

**2.7. Reliability as a Function of the Number of Replicates.** How does the reliability of gene classification vary with the number of replicates? For this experiment, a partial answer is provided by Table 5, which shows the percentages of the 288 genes that are misclassified by this methodology for each possible combination of one, two, and three replicates in the experiment. The false-positive and false-negative components of the misclassification percentage are also shown in the table. First, we note that false positives dominate. This result could be anticipated from our earlier findings and suggests that false indications of expression may be prevalent in microarray studies. Second, the table shows how classification precision varies with the number of replicates. A single replicate, such as replicate 2, may happen to have a low misclassification percentage (1.4%) relative to other replicates but, unfortunately, this reliability cannot be anticipated in advance. For example, replicate 3 alone misclassified 9.0% of genes. As expected, Table 5 confirms that average reliability and the certainty of that reliability increase with the number of replicates. We might surmise that the maximum attainable precision has been achieved with three replicates in our experiment, because the error rate appears to be leveling out at 0.7%. We note that there is no assurance the error rate will go to zero with increasing replication unless all sources of experimental variability are replicated, which is not the case in this experiment.

The optimal number of replicates in a general microarray study will depend on many factors, including the type of array equipment, laboratory technique, and the condition and preparation of samples. If experimental resources and time permit, we see potential benefit from using a minimum of three replicates because three or more classification outcomes offer the possibility of triangulation of results. A comparison of classification outcomes for all possible combinations of replicates, as is done for pairs of replicates in Table 3 *Right*, for example, might show whether one or more replicates are rogues. A judgment might then be made whether such replicates should be discarded. Replicates might also be used with a majority voting rule to decide whether a gene is expressed. Such a rule is not beneficial in this experiment but might be useful in some applications.

### Concluding Discussion

The findings of our simple experiment have three important implications for the generation, analysis, and interpretation of microarray data. First, we have shown that any single microarray output is subject to substantial variability even under the relatively controlled conditions of an experiment. By design, we have introduced only one potential source of variability, namely the location of spots on the slide. Variability from other sources, such as multiple preparations of probe, arrays on different slides, or arrays generated at different times, has not been admitted. Thus, our experiment is evaluating the minimum variability that is likely to be inherent in this system. Still the variation from this one source is considerable. A single output yields numerous misclassifications and, especially, numerous false positives. Replications of the experiment are not consistent and therefore produce different lists of expressed genes.

Table 5. Misclassification percentages for different combinations of replicates

Classification Outcome	Combination of Replicates						
	(1)	(2)	(3)	(1, 2)	(1, 3)	(2, 3)	(1, 2, 3)
False positive, %	8.3	1.4	9.0	1.0	2.1	0.7	0.7
False negative, %	0.3	0.0	0.0	0.3	0.3	0.0	0.0
Misclassified, %	8.7	1.4	9.0	1.4	2.4	0.7	0.7

Second, in modeling the random variation in gene expression, we have found in any single replicate the probability may be as large as 5% that mRNA in the sample tissue either fails to be represented as probe or, if it is represented as probe, fails to be hybridized to the cDNAs that are deposited on the slide (false negatives). Also, the probability may be as large as 10% that ghost genes are expressed (false positives). When microarray data from several replications are combined, we have shown that, quite reasonably, a more accurate genetic picture is produced with a reduction of false positives and false negatives. Third, in the process of analyzing these experimental data, we introduced statistical methodology for microarray data. We have modeled gene expression measurements by using a mixture of normal distributions. From this mixture distribution, a posterior probability is calculated from the microarray reading that quantifies the likelihood that the gene is truly expressed in the tissue. This probability can be used to classify whether a gene transcript is present. A two-way linear statistical model is proposed for microarray data that can span a range of experimental conditions.

Although our results depend on specific instruments and techniques (e.g., RNA extraction method, probe synthesis and labeling, hybridization, array construction, use of glass slide as solid support, and use of only one channel Cy3), the statistical methods we propose can be extended to accommodate more general settings. For example, the methods can be used for

experiments that use both channels Cy3 and Cy5. If the two-channel system is used in the standard way with mRNA from a test sample and a reference sample, differential gene expression becomes relevant (i.e., the interaction term of the two-way linear model). As there are then three states of expression (unexpressed, differentially expressed in favor of the test sample, and differentially expressed in favor of the reference sample), a three-component mixture model applies. The statistical methods also extend to data sets from experimental designs that involve additional sources of variability, such as variability introduced by multiple preparations of probes.

The main lesson to be learned from the study is that replication in microarray studies is not equivalent to duplication and hence is not a waste of scientific resources. Experimental replication is essential to reliable scientific discovery in genetic research. Understanding the sources of noise in the process, controlling it, and, if possible, eliminating it, are essential to drawing reliable inferences. By pooling data from replicates, we can provide a more reliable analysis of gene expression data.

We acknowledge with thanks the financial support provided for this research by National Institutes of Health grants HL40619-09 and EY12269-02 (M.-L.T.L.), and CA75354 (J.S.), by the National Foundation for Cancer Research (J.S.), by the Natural Sciences and Engineering Research Council of Canada (G.A.W.), and by the Social Sciences and Humanities Research Council of Canada (M.-L.T.L. and G.A.W.).

1. Eisen, M. (1999) *SCANALYZE User Manual* (Stanford Univ., Stanford, CA), Ver. 2.32.

2. Eisen, M. B., Spellman, P. T., Brown, P. O. & Bostein, D. (1998) *Proc. Natl. Acad. Sci. USA* 95, 14863-14868.

# Gene Expression Profile Analysis by DNA Microarrays

## Promise and Pitfalls

Hadley C. King, BA

Animesh A. Sinha, MD, PhD

**I**NCREASING NUMBERS OF HUMAN DISEASES, both acquired and genetic, are being considered to be based at least in part on alterations in DNA sequence. For most diseases, inheritance and acquisition are likely to be complex and polygenic. The efforts of the Human Genome Project to elucidate the structural genetic background by identifying the chromosomal positions and genomic organization of between approximately 30 000 and 35 000 human genes are nearly complete.<sup>1</sup> Based on this structural knowledge, a byproduct should be a better "scaffolding" to help link specific genes to susceptibility to various human diseases. However, to understand how the products of these genetic linkages work together to orchestrate the initiation and progression of particular complex diseases, there will be a need to apply a functional genetic rather than a structural genetic approach.<sup>2,3</sup>

Until recently, functional genetic studies have generally been of limited scope, only able to elucidate the role of 1 or a few genes at a time in 1 system. Information on the specificity and relative abundance of expression products has traditionally been obtained by techniques such as RNA Northern blot hybridization and ribonuclease protection assays. Somewhat more sophisticated methods, such as differential display<sup>4</sup> and Serial Analysis of Gene

DNA microarrays represent a technological intersection between biology and computers that enables gene expression analysis in human tissues on a genome-wide scale. This application can be expected to prove extremely valuable for the study of the genetic basis of complex diseases. Despite the enormous promise of this revolutionary technology, there are several issues and possible pitfalls that may undermine the authority of the microarray platform. We discuss some of the conceptual, practical, statistical, and logistical issues surrounding the use of microarrays for gene expression profiling. These issues include the imprecise definition of *normal* in expression comparisons; the cellular and subcellular heterogeneity of the tissues being studied; the difficulty in establishing the statistically valid comparability of arrays; the logistical logjam in analysis, presentation, and archiving of the vast quantities of data generated; and the need for confirmational studies that address the functional relevance of findings. Although several complicated issues must be resolved, the potential payoff remains large.

JAMA. 2001;286:2280-2288

www.jama.com

Expression,<sup>5</sup> have been used to screen larger numbers of complementary DNA (cDNA) clones. However, technical limitations render these techniques nonconductive to large-scale genetic survey.

To this end, a powerful new technology is emerging, using hybridization to nucleotide arrays, the so-called gene chips.<sup>6,7</sup> This technological intersection of biology and computers enables the reliable screening of a vast number of genes simultaneously and is amenable to automation. On a nylon membrane or glass surface, gene-specific cDNAs can be spotted, or oligonucleotides can be synthesized in situ by a combination of photolithography and oligonucleotide chemistry. This

permits simultaneous monitoring of the expression of thousands of genes in a single step. Individual chips can be customized to include any chosen set of fully or partially characterized genomic or expressed sequences. Chips can monitor over 50 000 unique sequences. The power of these chips lies in the potential for comparative expres-

**Author Affiliations:** Department of Dermatology, Weill Medical College of Cornell University, New York, NY. **Financial Disclosures:** Dr Sinha received reimbursement for material on hair loss prepared for a Healtology Web cast and received funding from Merck for research on microarrays involving alopecia.

**Corresponding Author and Reprints:** Animesh A. Sinha, MD, PhD, Department of Dermatology, Weill Medical College of Cornell University, 510 E 70th St, LC-702, New York, NY 10021 (e-mail: ans2003@med.cornell.edu).

**Toward Optimal Laboratory Use Section Editor:** David H. Mark, MD, MPH, Contributing Editor.

sion studies in diseased vs normal samples, and in documenting changes at different stages during the natural course of the disease or in response to treatment. It provides the researcher with a new arsenal to analyze underlying pathomechanisms on a grand scale and also to review the rationale of therapeutic concepts.

However, despite the enormous potential of this revolutionary technology, there are several issues and possible pitfalls that attenuate the power of microarrays. First, the definition of normal in expression comparisons is neither precise nor unambiguous. Second, the heterogeneity of the tissues being studied complicates the meaning of the expression profiles. Third, the statistically valid comparability of arrays is an unresolved problem. Fourth, the vast quantities of data create a logistical logjam for analysis, presentation, and archiving. Finally, confirmational studies are needed to corroborate the biological significance of microarray data (FIGURE 1).

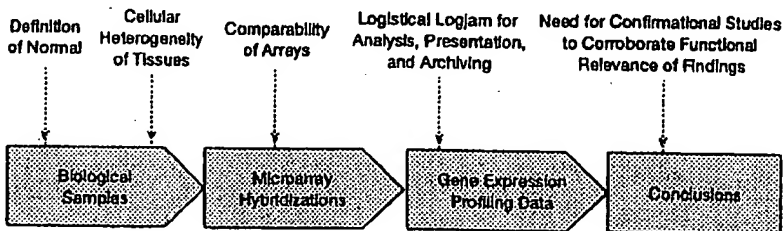
### TROUBLE WITH NORMAL

The standard normal vs diseased tissue type of comparison, which is the basic design foundation of profiling studies, may be more quicksand than bedrock. Normal is not so easy to define—neither is diseased. Gene expression in normal tissue is likely to be dependent on several factors involving patient and sample variation. These factors will also have an impact on expression profiles of diseased tissue.

#### Patient Variation: Ethnicity, Sex, Age, Genetic Background, Disease States

The ethnicity, sex, age, and genetic background of a patient are likely to affect the gene expression profiles of many tissues to varying extents.<sup>8,9</sup> A simple example is provided by the expression profiles of genes involved in scalp and body hair follicle activity, which can be expected to vary over a normal range under the influence of all of these sources of patient variation. The effects of these parameters on gene expression are likely

Figure 1. Possible Pitfalls in Microarray Gene Expression Profiling Analysis



There are issues that complicate each step of microarray gene expression analysis.

to be subtle but pervasive, not fully understood at this time, and quite problematic for defining normal.

The presence of disease in a subject who is the source of tissue for control purposes, presents further potential variabilities. For example, there may be a significant difference in the conclusions reached by 2 similar microarray expression profiling studies. One may compare genes expressed in a patient's diseased lung tissue with those expressed in normal, nondiseased lung tissue from the same patient, and another may compare genes expressed in the same patient's diseased lung tissue with those expressed in normal, nondiseased lung tissue from a healthy control or normal individual. Moreover, it is also possible that seemingly unrelated disease states may influence gene expression at distant sites. For instance, the presence of diabetes in 1 of 2 renal cancer patients may complicate the direct comparison of renal tissues.

#### Sample Variation: Proximity to Disease, Anatomic Location, and Developmental Range

Yet another complication derives from the proximity of the normal tissue used as a control for the diseased tissue. Tissue adjacent to an area of disease may not be normal despite absence of evidence of disease clinically or under the light microscope. Normal-appearing tissue near a tumor could, for example, be genotypically altered or exhibit an altered gene expression profile.<sup>10-13</sup>

Moreover, factors such as the degree of disease-associated inflammation may have a significant impact on gene expression profiles. Other bystander effects, epiphenomena, or secondary disease processes could all play important roles in determining expression profiles within these adjacent, so-called normal tissues. These factors must be considered in the choice of normal.

The precise location within a particular organ may be another important factor that affects gene expression.<sup>9</sup> For example, just as location relative to the urethra may influence expression profiles in the prostate,<sup>14</sup> skin from the nose, back, and palm are certain to have different expression profiles as well, despite all being from the same organ. Thus, site and specific anatomic location must also be taken into account in a description of normal.

It must also be kept in mind that the definition of normal actually represents a dynamic state.<sup>14</sup> All tissues, which are composed of early and late-stage cells, have a normal developmental range. For example, normal epithelium in prostatic ducts ranges from atrophic to resting to hyperplastic, and each has a unique pattern of gene expression.<sup>14</sup>

A 3-dimensional analytic approach is a strategy that has been used to address some of these concerns about defining normal. Cole et al<sup>14</sup> used a 3-dimensional model to characterize the entire prostate gland in their study of gene expression profiles in prostate cancer. In this study, whole-mount prostatectec-



tomy specimens were divided into transverse cross sections such that the entire prostate gland, including the complete spectrum of normal epithelium and tumor progression, was available for viewing, microdissection, and microarray analysis. This method was used to determine the exact physical relationship of the normal ducts, premalignant lesions, and tumors—thus obtaining an anatomic framework on which to overlay gene expression data. This technique offers several advantages over the normal vs tumor comparison. Previous studies had used normal epithelium in prostatic ducts as a baseline control against which to compare and contrast tumor gene expression profiles.<sup>13,16</sup> However, the expression profile of this normal epithelium is affected by proximity to tumor, location within the gland, and developmental state.<sup>14</sup> These factors can be better appreciated using a 3-dimensional approach.

#### Disease-Related Variation

Of course, many of the parameters that affect normal expression profiles (patient ethnicity, age, sex, and genetic background, location within an organ, and developmental stage) will also affect disease expression profiles.<sup>17-19</sup> Disease heterogeneity, including subtype, activity, severity, stage, and previous as well as current treatments, also may have a significant impact on gene expression.<sup>20-23</sup>

Categorizing and subgrouping patients on entry into a study may be useful to control for as many of these factors as possible. However, there may be problems surrounding attempts to define microarray-based categorization on the basis of another imperfect categorization system, such as histology, as these groups are sometimes arbitrary or inconsistently designated. Nevertheless, determining whether gene expression profiles correlate with existing clinical or histological categories can provide new insights into the meaning of these categories as can new methods of classifying cancers or other diseases into specific diagnostic categories based on their gene expression signatures. Several stud-

ies have been able to establish expression-based criteria (class predictors) for pre-existing categories and then use these new criteria to categorize new cases (class prediction).<sup>26-28</sup> Global profiling may also allow the development of new classification systems based on gene expression alone (class discovery).<sup>29,30</sup>

Thus, when possible, it will be of value to profile a range of normal and diseased cell populations from a number of patients to distinguish between differences in expression that are relevant to the disease process and those reflecting the biological spectrum of the normal tissue or that have occurred for reasons unrelated to the disease. The significance of this distinction is further appreciated when taking into account the vast quantity of data generated from microarrays and the potential for confounding interpretation from the inclusion of differential expression unrelated to disease processes.

It is worth noting, however, that the issues of patient and sample variability are not unique to microarray experiments. In fact, microarray experiments, in contrast to classic single-gene experiments, may actually provide the tools for identifying this heterogeneity. For example, DNA microarrays have been used to explore physiological variation in gene expression on a genomic scale in 60 cell lines derived from diverse tumor tissues.<sup>31</sup> Cluster analysis allows the identification of prominent features in gene expression patterns that appear to reflect molecular signatures of the tissue from which the cells originated.<sup>31</sup>

#### HETEROGENEOUS CELL POPULATIONS

A further complication encountered with expression profiles is that any given tissue is composed of several cell types, members of which are likely to be within a spectrum of dynamic functional states. For example, a simple punch biopsy of the skin may contain keratinocytes, melanocytes, Langerhans cells, Merkel cells, adipocytes, smooth muscle cells of arrector pili, striated muscle cells of the panniculus carnosus, blood cells including immune system cells, and cellular el-

ements of blood vessels, nerves, hair follicles, sebaceous glands, and sweat glands. Moreover, cells from each of these populations will be at various stages of development and levels of activation, performing different functions and responding to disease processes or treatments in different ways and to varying extents. The result is a highly heterogeneous sampling of cells, each expressing a unique set of genes. An expression profile generated from a microarray study of the RNA in such a sample will thus represent merely a snapshot of the genes expressed by a plethora of cells at a moment in time. Such extensive cellular heterogeneity complicates the ability to draw conclusions about specific processes occurring within a tissue specimen. An illustrative example is provided by Stanton et al,<sup>32</sup> who used microarrays to identify genes differentially expressed during myocardial infarction. The expression profiles they studied represented transcripts from cell populations as diverse as immune system cells, which migrated to the infarct region and are responsible for the inflammatory response, cardiac myocytes within the ischemic area undergoing apoptosis and necrosis, fibroblasts undergoing proliferation and participating in the formation of scar tissue to replace the infarct, and cardiac myocytes undergoing hypertrophy to compensate for the loss of cells in the infarct area.<sup>33</sup> The issue of such cellular heterogeneity was avoided by categorizing the differentially expressed genes into functional categories to look for patterns indicative of cardiac remodeling without attempting to attribute specific transcripts to specific cell types. For gene expression studies involving samples with mixed cellular populations, further investigation, such as with *in situ* messenger RNA (mRNA) hybridization, may be necessary to localize the transcripts before conclusions can be drawn about the roles of specific genes in specific cell types during the disease process.

#### Laser Capture Microdissection

An ingenious but technically delicate approach to the study of complex bio-

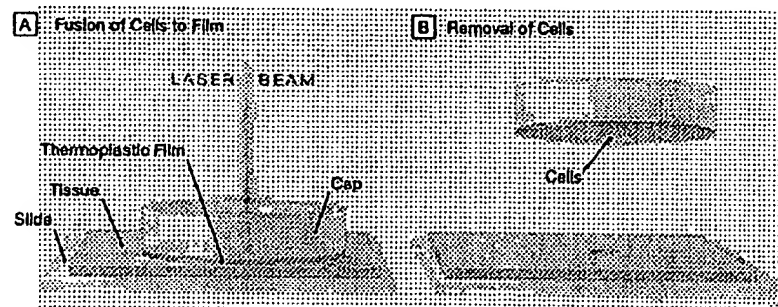


logical samples has become possible with the development of laser capture microdissection (FIGURE 2).<sup>34</sup> This technique allows for the rapid and accurate procurement of cells from specific areas of tissue under direct microscopic visualization, and thus makes the molecular genetic analysis of defined populations in their native tissue environment possible.<sup>35</sup> Sgroi et al<sup>36</sup> demonstrated the feasibility of combining laser capture microdissection with high-throughput cDNA arrays. They showed that *in vivo* subpopulations of malignant cells from multiple stages of breast cancer progression could be separated from nonmalignant populations, and their expression profiles could subsequently be analyzed using microarrays.

The potential is powerful. Specimens could be separated into tissue layers; for example, separating a skin biopsy into epidermis, dermis, and hypodermis. Tissues could be further differentiated into specific structural components, such as dermis into blood vessels, adipose, arrector pili, and sebaceous glands. Structures could be separated into defined cell types, such as blood vessels into endothelial cells, erythrocytes, and lymphocytes. Cell types could even be separated into marker-defined subtypes, such as lymphocytes into CD4 and CD8 cells. Expression profiles from refined and defined structures and cell types likely would be extremely valuable in the study of disease.

Potential aside, there are significant limitations to this technology at the present time. The standard protocols for fixing and embedding tissue samples from surgical resections were not designed to be compatible with microarray experiments, with or without laser capture microdissection. Typically, tissue suspected of being important for diagnosis and staging is processed through aldehyde-based fixatives, such as formalin, which damage mRNA integrity.<sup>37</sup> If frozen tissue is available, mRNA can be recovered and studied from dissected cell populations. However, frozen tissue sections are technically difficult to prepare, the histology is often

Figure 2. Laser Capture Microdissection



A, Tissue sections are processed and placed on a microscope slide under a thin, transparent thermoplastic film, which is attached to a movable cap. Visualizing the tissue microscopically, a short-duration, focused pulse from an infrared carbon-dioxide laser is used to activate and melt the film to selectively adhere cells within targeted areas of interest. B, When the cap is lifted, the film, with selected cells still bound, is removed from the tissue section for further processing to retrieve cellular materials (eg, DNA, RNA, proteins).

severely compromised, and the tissue available may contain only a limited portion of the lesion.<sup>14</sup>

Moreover, the sample amounts generated from laser capture microdissection can be small, even as miniscule as a single cell.<sup>38</sup> Consequently, the yields of RNA are low. Arrays have a threshold for the quantity of molecular starting material: at least 5 to 15 µg for oligonucleotide arrays and between 2 and 100 µg for cDNA arrays, depending on the manufacturer, the source of the RNA, and the use of signal amplification.<sup>39,40</sup> Studies that have successfully integrated laser capture microdissection with microarray technology have used samples of approximately  $1 \times 10^4$  to  $1 \times 10^5$  cells with 95% to 98% homogeneity as determined by microscopic visualization.<sup>36,41</sup> If needed, amplification techniques may be used to generate sufficient genetic material for microarray hybridizations.<sup>41</sup> Laser capture microdissection is an intriguing technology, but time will tell whether its potential is realized.

Although some biological issues related to gene expression may be complicated by the presence of heterogeneous cell populations in studied samples, it is also true that some biological conditions can be understood only in the context of these heterogeneous cell populations. The nature of

global gene expression experiments is to uncover differences between 2 biological samples, including those differences based on diverse cell populations. For example, to appreciate a disease that is characterized by an inflammatory infiltrate, it must be understood that the inflammatory infiltrate is part of the disease and is part of the difference between diseased and nondiseased tissue. Thus, the isolation of specific cell populations for study is not necessarily required or even desirable in all instances.

## MAKING MICROARRAYS COMPARABLE

Ideally, microarray experiments should be comparable both within and between laboratory or manufacturing systems, but obtaining consistent and comparable data is a critical challenge for microarray-based expression analysis. Major sources for the observed variability of microarray data include the normal physiological gene expression variations in different samples and the noise introduced in the microarray assay process.<sup>42</sup>

## Physiological Gene Expression Variation

Inextricably linked to the issues of patient and sample variability and tissue heterogeneity discussed above, is the

problem of normal gene expression variations and how to distinguish these variations from significant disease-associated changes. Few studies have systematically investigated physiological expression changes, but data from *in situ* hybridizations suggest that normal variance for many tightly regulated tissue-specific genes can be within 20% to 30%.<sup>42</sup> However, there can be as much as 2- to 4-fold random fluctuations for many genes in yeast.<sup>43,44</sup> Affymetrix (Santa Clara, Calif) guidelines have suggested that for most of the "housekeeping" genes in human tissues, which are likely to be less tightly regulated, differences of less than 4-fold are probably not biologically significant.<sup>45</sup> Consequently, a significant portion of microarray data variability for high- or medium-abundance mRNAs may simply be due to normal expression variations. Several previous studies have used arbitrary 2-fold change criteria to define significant expression change.<sup>46</sup> However, the 2-fold threshold has been shown to be statistically invalid even for duplicate experiments.<sup>46</sup> In a recent study that used cDNA microarrays to profile gene expression in samples of normal skin from breast-reduction surgeries, 71 of 4400 genes were found to demonstrate variability in expression greater than 2 SDs from the mean of each gene.<sup>47</sup> These included genes coding for transport proteins, gene transcription, cell-signaling proteins, and cell-surface proteins. Thus, physiological variation should be considered in the analysis and interpretation of microarray data. More stringent criteria for defining significant expression change may be useful.

#### Noise in the Microarray Assay Process

For the tightly regulated (mostly low abundance) mRNA species, inconsistencies introduced at any stage of the microarray-based assay process may play a major role in data variability.<sup>42</sup> Due to the miniaturization and the large number of genes involved, it is difficult to maintain consistent processing conditions for each sequence across multiple assays, and obtaining accu-

rate absolute signals is unlikely.<sup>42</sup> Noise may be introduced by slide heterogeneities, printing irregularities (eg, pin-to-pin variations), and spotting volume fluctuations.<sup>48</sup> Some of the systematic variations may be reduced by the inclusion of controls, but random fluctuation at various manufacturing stages cannot be completely controlled and can accumulate quickly in a complicated assay.<sup>48</sup>

In certain microarray systems, 2 samples are competitively hybridized to 1 array using different fluorors for labeling. In other systems, there is only single-sample hybridization. A 2-color system might be expected to be more reliable since variations in spot size or amount of cDNA probe on the chip should not affect the signal ratio (both signals are derived from the same spot). However, this only holds true if signals are well above the background in both detection channels.<sup>42</sup> In fact, the signal level for most of the tightly regulated genes will likely be close to the background level.<sup>42</sup> In addition, background level on a slide can also vary significantly from spot to spot due to factors such as unevenness in slide surface properties, dust contamination, and incomplete washing, leading to high levels of signal variability for low-abundance mRNA species even in 2-color systems.<sup>42</sup>

The high levels of variability of microarray data also mean that subtle changes in experimental conditions may significantly alter the results, making it difficult for separate laboratories to compare experimental data. In addition, the lack of standard controls, the predominant use of relative signals (ratios), and the adoption of incompatible data formats contribute to poor comparability between studies.<sup>42</sup>

Despite the hard-wired variability introduced by chip manufacturing conditions, most of the published studies to date using microarray-based expression analysis include only limited numbers of replicates.<sup>49</sup> In fact, many studies conduct the experiment only once. Considering the potential sources of assay variation, the need for sufficiently replicated studies is underscored.<sup>49</sup>

#### Microarray Data Normalization

Because of variability of microarray data for single sample arrays and for further analysis of 2-color system arrays, each must be brought into the same scale to compare 2 or more arrays. How to perform this normalization of gene expression levels across multiple arrays, thus removing systematic variation between the arrays and rendering different experiments comparable, remains an issue that is not yet fully resolved.<sup>50</sup> Many of the early microarray studies in the literature simply ignored this issue. A more statistically rigorous approach is needed.

One difficulty has been that leading microarray manufacturers have not published statistical error models for their products. Thus, users are unclear how much to adjust data for variations in spot intensity, hybridization patterns, and intensity measurement sensitivity. Software does exist to allow for array-to-array comparisons by using a scaling factor to normalize gene expression patterns across arrays. However, in general, these algorithms assume that intensity differences between arrays are linearly related with a zero y-intercept.<sup>51</sup> This assumption allows software to trim the tails off distributions of expression from different arrays at statistical cutoffs and then simply move the distributions along an axis to a common level to provide comparisons. However, this linear relationship often does not hold true.<sup>51</sup> When the average expression level of 1 array is higher than that of a second array, a longer tail will be trimmed from the second array. Thus, a greater number of genes from the first array will be counted as being expressed because their expression level is above the statistical cutoff point. In this case, the 2 arrays cannot be considered comparable.

Although bioinformatic software has recently been developed that offers more statistically robust normalization, the cost of these commercially available programs (combined with the already expensive microarrays) has been prohibitive for many research-

ers.<sup>50</sup> Standardization of these processes awaits the development of improved methods of normalization leading to valid statistical models widely available to all researchers.

To this end, Schadt et al<sup>51</sup> have developed a standard nonlinear curve technique for normalizing the data in arrays that do not demonstrate a linear relationship between data sets. This model performs well when the 2 samples being compared demonstrate a low number of differentially expressed genes. However, when expression profiles of 2 samples vary to a greater extent, Schadt et al<sup>51</sup> recommend a rank-selection method. Using this method, genes expressed on an array are ranked from highest to lowest level of expression. Then, for the array expressing a greater number of genes, the genes with the lowest expression levels are removed from the list until the 2 arrays list a comparable number of expressed genes. This type of rank-selection method has gained support from other groups, but it too has limitations.<sup>50</sup> Removing low-expression level data points restricts the study to the more extreme and easily detected entities, a technique that blunts the genomic-scale potential of microarray technology.

Efforts continue to improve comparability between arrays. Jones<sup>50</sup> recently applied a statistical model to normalize spotted cDNA array data that takes into account not only the differences in numbers of genes expressed between arrays, but also the interarray variations in fluorescent dye intensity and mechanical error occurring in the printing process. Nevertheless, the issue of how to properly normalize array data has not been settled. Researchers must continue to demand statistical rigor in their comparisons before they can believe the mathematical results of their data.

### LOGISTICAL LOGJAM

Microarrays deliver massive amounts of data on tens of thousands of genes. The result is an immense quantity of biological information that must be ana-

lyzed, presented, and archived in a meaningful way.

### Data Analysis

In human studies, the number of hybridizations that can be performed for any set of experimental conditions is often restricted by the limited number of obtainable tissue samples and by the expense of arrays. Restricted numbers of hybridizations for each experiment hamper the ability to assess the biological significance of variation within or between given sets of conditions. Thus, for the assessment of thousands of genes in a setting of limited hybridizations, the importance of reliable and sophisticated algorithms for data analysis becomes amplified.<sup>51</sup>

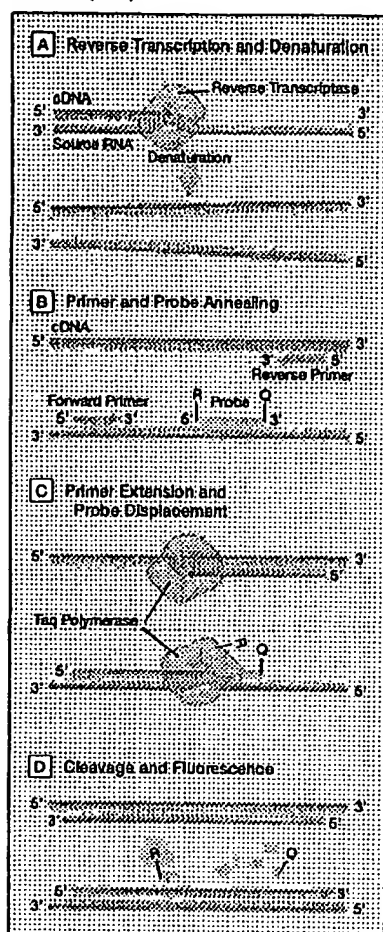
A logical beginning is to examine the extremes, that is, genes with significant differential expression in individual samples. For example, a comparison of 2 samples can be visualized in the form of a simple bivariate scatterplot in which the expression profile of 1 sample (x-axis) is plotted against that of the second sample (y-axis). The distribution pattern generally demonstrates that the expression ratios cluster around the line in which x is equal to y (indicating comparable levels), with individual genes falling varying distances from this line. Additional lines can be placed on the scatterplot to represent various fold changes of expression. Data points that fall above or below these lines represent genes exhibiting expression ratios greater or less than the specified fold change. Thus, one can begin by examining those genes that demonstrate a 10-fold or greater change in expression level. To expand the number of genes under investigation, one can examine genes that demonstrate a 5-fold or greater change, or a 3-fold or greater change, and so forth. Many studies define a 2-fold or greater change in expression level to represent significant differential expression. The 2-fold threshold, however, as noted above, has been shown to be statistically invalid.<sup>46</sup> Although this simple technique can be efficient and effective for focusing on expanding sets of differentially expressed

sequences, again, such an analysis does not take advantage of the full potential of genome-scale experiments to enhance our comprehension of cellular biology that would be provided by an inclusive analysis of the entire repertoire of transcripts in a cell as it goes through a biological process.<sup>52</sup> A more holistic approach, which allows the deciphering of patterns from the entire data set, is needed.

### Data Organization and Presentation

Statistical algorithms can be applied to detect and extract patterns within profiling data. It is a basic assumption of many gene expression studies that knowledge of where and when a gene is expressed provides information about the function of the gene. Therefore, an important beginning is to organize genes on the basis of similarities in their expression profiles.<sup>53</sup> However, even this basic tenet deserves critical consideration. Similarity of gene expression profile does not mandate similarity of function or mechanistic pathway, and it may be purely coincidental. Nevertheless, the idea of clustering genes on the basis of their expression patterns is well established and cluster analysis has become the most widely used statistical technique applied to large-scale gene expression data.<sup>52</sup>

Although various cluster methods can usefully organize tables of gene expression measurements, the resulting ordered but still massive collection of numbers remains difficult to assimilate. Thus, another important component of genome-wide expression data exploration is the development of powerful data visualization methods and tools. Approaches have been developed that present clustering results in simple graphical displays such as dendrograms, which represent relationships among genes by a tree whose branch lengths reflect the degree of similarity in expression between the genes. Similarity is mathematically defined.<sup>54</sup> The computed trees can be used to order genes in the original data table such that genes or groups of genes with similar expression level patterns are

**Figure 3.** Real-Time Polymerase Chain Reaction (PCR)

A, Complementary DNA synthesis from source messenger or total RNA proceeds as with traditional reverse transcriptase PCR. B, A dual-labeled fluorogenic probe (with a higher melting point than the PCR primers used for extension) is annealed to the target sequence between the forward and reverse PCR primers. A reporter (R) fluorochrome (usually 6-carboxy-fluorescein) is attached to the 5' end of the probe and a quencher (Q) molecule (usually 6-carboxy-tetramethyl-rhodamine) is attached to the 3' end. C, As Taq polymerase extends in the 5' to 3' direction, the dual-labeled probe begins to be displaced from the target sequence. D, As the Taq polymerase continues to extend, the 5' to 3' endonuclease activity cleaves the reporter (R) molecule from the probe sequence such that its emission spectra (518 nm) are no longer quenched by the second (Q) fluorescent dye. Fluorescence is measured continuously throughout the PCR amplification in real time and is proportional to the amount of PCR product generated in each cycle.

placed adjacently. Clustering methods can also be combined with representation of each data point with a color that quantitatively and qualitatively reflects

the original experimental observations.<sup>52</sup> Visual assimilation is then more intuitive.

### Data Archiving and Mining

Ultimately, successful interpretation of gene profiling studies is likely to be dependent on the integration of experimental data with external information resources. As multiple experiments involving multiple cell types and tissues from multiple laboratory groups accumulate, data archiving may well become the watershed issue. Ideally, all data, in a suitably standardized form, would be freely accessible in the public domain. Even assuming a willingness to share the data, such utopian goals would require a user-friendly and powerful database system and standardization of correction and normalization procedures such that data points from various projects become comparable.<sup>55</sup> The National Center for Biotechnology Information Entrez system (<http://www.ncbi.nlm.nih.gov/Entrez/>) does provide useful data in this regard, but current databases may be limited in scope or computability.<sup>53</sup> A major focus of infrastructure development to support genomic-scale expression studies will need to be in the area of electronic biological pathway databases and resources.

### CONFIRMATIONAL STUDIES

The development of more sophisticated analytical algorithms and databases will help lend credence to the biological significance of differential gene expression determined by microarray analysis. In the meantime, several studies have begun to examine the sensitivity and specificity of microarray-based experiments. Sensitivity, defined as the minimum reproducible signal detected by a given array scanning system, has been reported for microarrays to be approximately 10 mRNA copies per cell, which is slightly inferior to the sensitivity of Northern blot analyses.<sup>56,57</sup> Specificity studies showed that for a given probe any nontarget transcripts with more than 75% sequence similarity may show cross-

hybridization.<sup>56</sup> The problem of clone misidentification and the need for clone confirmation have also been addressed.<sup>58</sup> One study found that of 1189 bacterial stock cultures, only 62.2% were uncontaminated and contained cDNA inserts that had significant sequence identity with published data for the ordered clones.<sup>59</sup> Thus, the use of sequence-verified clones for cDNA microarray construction is warranted.

Additionally, potential gene candidates can be assessed for relevance to disease using parallel technologies. Several such alternative platforms have been used to bolster the importance of specific sequences first suggested in gene chip comparisons including (1) methods at the RNA level, (2) methods at the protein level, and (3) functional studies.

### Methods at the RNA Level

Reverse transcriptase polymerase chain reaction (RT-PCR) is a method often used to verify microarray data. Although RT-PCR is not well suited to quantitation, the relative technological ease of this assay and the ability to rapidly monitor multiple samples make it a useful technology.<sup>60,61</sup> Hybridization data can be verified and multiple putative markers can be screened in a short period.

Several other studies have used real-time quantitative RT-PCR (TaqMan, PE Applied Biosystems, Foster City, Calif).<sup>15,62</sup> Real-time PCR is a technique that increases the quantitative ability of RT-PCR by providing accurate and reproducible information on RNA copy number (FIGURE 3). In this method, a fluorogenic probe (labeled at the 5' end with a reporter fluorochrome and at the 3' end with a quencher fluorochrome) is annealed to 1 strand of the target cDNA sequence between the forward and reverse PCR primers. As Taq polymerase extends the forward primer, its intrinsic 5' to 3' nuclease activity displaces and degrades the dual-labeled probe, releasing the reporter fluorochrome from the quencher label and allowing the detection of a fluorescent

signal that is proportional to the amount of PCR product generated in each cycle.<sup>63</sup>

Northern blot analysis is also commonly used as a confirmational technique, as it is a standard specific and semiquantitative method.<sup>15,57,61</sup> For mRNA expressed at moderate-to-high levels, and for which cDNA probes are available, Northern blot analysis works well, but it is not well suited for low-copy mRNA.<sup>64,65</sup> Furthermore, only a small number of genes can be analyzed with this conventional method.

#### Methods at the Protein Level

DNA microarray technology is limited to the study of gene expression at the mRNA level. However, it has been established that mRNA levels do not necessarily correlate with protein levels. Moreover, the level of expression or even presence of a protein is not tightly linked to physiological consequences. An investigation conducted by Winzeler et al,<sup>66</sup> for example, provides a cautionary tale. Their study demonstrated that genes up-regulated in yeast growing in minimal medium did not prove to be more important for growth than genes that were not upregulated.<sup>39</sup> They found only 2 of 8 genes required for yeast growth in minimal medium to be induced. The lesson to be learned is that genes that are not differentially expressed may be of equal functional importance in disease states compared with those that are.

Furthermore, the regulation of some genes may be at the translational rather than the transcriptional level, which would preclude detection by DNA microarrays. Posttranslational modification of proteins is also an important mode of regulation that cannot be detected by DNA microarrays. Protein activity, particularly receptor activity, is heavily dependent on phosphorylation, for example. DNA and mRNA reveal nothing about whether a given protein is active, and can be deceptive when used to speculate about quantities of proteins. It has been demonstrated that the correlation between mRNA and protein abundance is less than 0.5,<sup>67</sup> emphasizing that ideally, mRNA expres-

sion studies should be accompanied by analyses at the protein level.<sup>39</sup> Radioimmunoassay and immunohistochemistry have been used in a number of studies.<sup>15,68,69</sup> These techniques, however, are not well suited to detecting low levels of expression, and they require the availability of an antibody specific for the protein to be studied.

The field of proteomics, the large-scale parallel analysis of the proteins that are present in a cell, is developing rapidly, but has problems of its own. Proteins vary in abundance by many orders of magnitude within a given cell, and there is no PCR equivalent for the amplification of proteins. Moreover, proteins fold in many known (and unknown) ways that affect their function. The feasibility of the microarray analysis of proteins has begun to be explored. Antibodies attached to microarrays can be used to bind to and quantitatively detect proteins that have been tagged with fluorescent dyes.<sup>70</sup> Skeptics doubt the plausibility of identifying thousands of unknown proteins in this manner.<sup>70</sup> The diverse chemistry of various proteins poses serious difficulties, and it will be challenging to find antibodies for every protein. Thus, although it is important to incorporate protein analyses into expression profiling studies, current platforms are technically limiting.

#### Functional Studies

Confirming the role of a gene initially identified in a microarray experiment in animal models with transgene or knockout studies provides a particularly powerful alternative platform. Transcript function, rather than mere presence, is addressed. However, this approach is ill-suited for high-throughput conditions. It may be ideal for an in-depth investigation of 1 or 2 genes of interest, but it is not practical for confirming large quantities of profiling data.

Confirmational studies are useful to corroborate the biological significance of differential gene expression determined by microarray analysis. While improved databases and more reliable statistical models will help to lend greater authority to array data, alter-

native platforms can be used to assess the relevance of genes first identified by array comparisons. It should be realized, however, that the alternative technologies are not intended for large-scale analyses. Realistically, only selected sequences from the array data can be confirmed with other platforms in the short-term, a retreat from the initial purpose of the genome-scale investigation by microarray.

#### CONCLUSIONS

Microarrays can be expected to prove extremely valuable as tools for the study of the genetic basis of complex diseases. The ability to measure expression profiles across entire genomes provides a level of information not previously attainable. Although complicated issues must be resolved, the potential payoff is big. Microarrays make it possible to investigate differential gene expression in normal vs diseased tissue, in treated vs nontreated tissue, and in different stages during the natural course of a disease, all on a genomic scale. Gene expression profiles may help to unlock the molecular basis of phenotype, response to treatment, and heterogeneity of disease. They may also help to define patterns of expression that will aid in diagnosis as well as define susceptibility loci that may lead to the identification of individuals at risk. Finally, as specific genes are identified and their functional roles in the development and course of disease are characterized, new targets for therapy should be identified.

Despite the problems of defining normal, understanding tissue heterogeneity, making arrays comparable, analyzing and archiving massive quantities of data, and performing confirmational studies in alternative platforms, expression profiling with microarrays stands as a truly revolutionary technology. As we continue to delve into the possibilities, we will surely progress in our understanding of current issues and complications. No doubt the ride on the high-throughput highway will be exhilarating.

*Author Contribution:* Ms King and Dr Sinha both contributed to the conceptualization and writing of this article.

## REFERENCES

1. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. *Science*. 2001;291:1304-1351.
2. Fields S. The future is function. *Nat Genet*. 1997;15:325-327.
3. Lander ES. The new genomics. *Science*. 1996;274:536-539.
4. Liang P, Pardee AB. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science*. 1992;257:967-971.
5. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. *Science*. 1995;270:484-487.
6. Lockhart DJ, Dong H, Byrne MC, et al. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol*. 1996;14:1675-1680.
7. Strachan T, Abitbol M, Davidson D, Beckmann JS. A new dimension for the human genome project. *Nat Genet*. 1997;16:126-132.
8. Nishimoto IN, Hanaoka T, Sugimura H, et al. Cytochrome P450 2E1 polymorphism in gastric cancer in Brazil. *Cancer Epidemiol Biomarkers Prev*. 2000;9:675-680.
9. Furuya KN, Gebhardt R, Schuetz EG, Schuetz JD. Isolation of rat pgp3 cDNA. *Biochim Biophys Acta*. 1994;1219:636-644.
10. Deng G, Lu Y, Zlotnikov G, et al. Loss of heterozygosity in normal tissue adjacent to breast carcinomas. *Science*. 1996;274:2057-2059.
11. Zhuang Z, Vortmeyer AO, Mark EJ, et al. Barrett's esophagus. *Cancer Res*. 1996;56:1961-1964.
12. Hung J, Kishimoto Y, Sugio K, et al. Allele-specific chromosome 3p deletions occur at an early stage in the pathogenesis of lung carcinoma [published correction appears in *JAMA*. 1995;273:1908]. *JAMA*. 1995;273:558-563.
13. Shimada S, Shiomori K, Tashima S, et al. Frequent p53 mutation in brain (fetal)-type glycogen phosphorylase positive foci adjacent to human 'de novo' colorectal carcinomas. *Br J Cancer*. 2001;84:1497-1504.
14. Cole KA, Kitzman DB, Emmert-Buck MR. The genetics of cancer—a 3D model. *Nat Genet*. 1999;21:38-41.
15. Xu J, Stolk JA, Zhang X, et al. Identification of differentially expressed genes in human prostate cancer using subtraction and microarray. *Cancer Res*. 2000;60:1677-1682.
16. Elek J, Park KH, Narayanan R. Microarray-based expression profiling in prostate tumors. *In Vivo*. 2000;14:173-182.
17. Chia SJ, Tang WY, Elnatan J, et al. Prostate tumors from an Asian population. *Br J Cancer*. 2000;83:761-768.
18. Dong M, Nio Y, Tamura K, et al. K-ras point mutation and p53 expression in human pancreatic cancer. *Cancer Epidemiol Biomarkers Prev*. 2000;9:279-284.
19. Pettaway CA. Racial differences in the androgen/androgen receptor pathway in prostate cancer. *J Natl Med Assoc*. 1999;91:653-660.
20. Hata K, Fujiwaki R, Nakayama K, Miyazaki K. Expression of the endostatin gene in epithelial ovarian cancer. *Clin Cancer Res*. 2001;7:2405-2409.
21. Salda T. Recent advances in melanoma research. *J Dermatol Sci*. 2001;26:1-13.
22. Hegde U, Wilson WH. Gene expression profiling of lymphomas. *Curr Oncol Rep*. 2001;3:243-249.
23. Liu L, Yang K. A study on C-erbB2, nm23 and p53 expressions in epithelial ovarian cancer and their clinical significance [in Chinese]. *Zhonghua Fu Chan Ke Za Zhi*. 1999;34:101-104.
24. Zhang Z, Dubois RN. Detection of differentially expressed genes in human colon carcinoma cells treated with a selective COX-2 inhibitor. *Oncogene*. 2001;20:4450-4456.
25. Oguri T, Isobe T, Fujitaka K, et al. Association between expression of the MRP3 gene and exposure to platinum drugs in lung cancer. *Int J Cancer*. 2001;93:584-589.
26. Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer. *Science*. 1999;286:531-537.
27. Khan J, Wei JS, Ringner M, et al. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med*. 2001;7:673-679.
28. Zhang H, Yu CY, Singer B, Xiong M. Recursive partitioning for tumor classification with gene expression microarray data. *Proc Natl Acad Sci U S A*. 2001;98:6730-6735.
29. Bittner M, Meltzer P, Chen Y, et al. Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature*. 2000;406:536-540.
30. Welsh JB, Zarrinkar PP, Sapinoso LM, et al. Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer. *Proc Natl Acad Sci U S A*. 2001;98:1176-1181.
31. Ross DT, Scherf U, Eisen MB, et al. Systematic variation in gene expression patterns in human cancer cell lines. *Nat Genet*. 2000;24:227-235.
32. Stanton LW, Garrard LJ, Damm D, et al. Altered patterns of gene expression in response to myocardial infarction. *Circ Res*. 2000;86:939-945.
33. Abdellatif M. Leading the way using microarray. *Circ Res*. 2000;86:919-920.
34. Bonner RF, Emmert-Buck M, Cole K, et al. Laser capture microdissection. *Science*. 1997;278:1481-1483.
35. Emmert-Buck MR, Bonner RF, Smith PD, et al. Laser capture microdissection. *Science*. 1996;274:998-1001.
36. Sgrol DC, Teng S, Robinson G, et al. In vivo gene expression profile analysis of human breast cancer progression. *Cancer Res*. 1999;59:5656-5661.
37. Klimeck WT, Futscher BW, Dalton WS. Effects of ethanol and paraformaldehyde on RNA yield and quality. *Biotechniques*. 1994;16:1021-1023.
38. Dolter KE, Braman JC. Small-sample total RNA purification. *Biotechniques*. 2001;30:1358-1361.
39. van Hal NL, Vorst O, van Houwelingen AM, et al. The application of DNA microarrays in gene expression analysis. *J Biotechnol*. 2000;78:271-280.
40. Burgess JK. Gene expression studies using microarrays. *Clin Exp Pharmacol Physiol*. 2001;28:321-328.
41. Kitahara O, Furukawa Y, Tanaka T, et al. Alterations of gene expression during colorectal carcinogenesis revealed by cDNA microarrays after laser capture microdissection of tumor tissues and normal epithelia. *Cancer Res*. 2001;61:3544-3549.
42. Watson SJ, Meng F, Thompson RC, Akl H. The "chip" as a specific genetic tool. *Biol Psychiatry*. 2000;48:1147-1156.
43. Cho RJ, Campbell MJ, Winzler EA, et al. A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol Cell*. 1998;2:65-73.
44. Klevecz RR, Kauffman SA, Shymko RM. Cellular clocks and oscillators. *Int Rev Cytol*. 1984;86:97-128.
45. Warrington JA, Nair A, Mahadevappa M, Tsyganskaya M. Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. *Physiol Genomics*. 2000;2:143-147.
46. Claverie JM. Computational methods for the identification of differential and coordinated gene expression. *Hum Mol Genet*. 1999;8:1821-1832.
47. Cole J, Tsou R, Wallace K, et al. Comparison of normal human skin gene expression using cDNA microarrays. *Wound Repair Regen*. 2001;9:77-85.
48. Schuchhardt J, Beule D, Malik A, et al. Normalization strategies for cDNA microarrays. *Nucleic Acids Res*. 2000;28:E47.
49. Lee ML, Kuo FC, Whitmore GA, Sklar J. Importance of replication in microarray gene expression studies. *Proc Natl Acad Sci U S A*. 2000;97:9834-9839.
50. Jones MM. Researchers attempt to defuse the microarray data minefield. Available at: <http://www.genomeweb.com/articles/view.asp?Article=200142175122>. Accessibility verified October 8, 2001.
51. Schadt EE, U C, Su C, Wong WH. Analyzing high-density oligonucleotide gene expression array data. *J Cell Biochem*. 2000;80:192-202.
52. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A*. 1998;95:14863-14868.
53. Bassett DE Jr, Eisen MB, Boguski MS. Gene expression informatics—It's all in your mine. *Nat Genet*. 1999;21:51-55.
54. Alon U, Barkai N, Notterman DA, et al. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc Natl Acad Sci U S A*. 1999;96:6745-6750.
55. Granjeaud S, Bertucci F, Jordan BR. Expression profiling. *Bioessays*. 1999;21:781-790.
56. Kane MD, Jatkoe TA, Stumpf CR, et al. Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nucleic Acids Res*. 2000;28:4552-4557.
57. Taniguchi M, Miura K, Iwao H, Yamanaka S. Quantitative assessment of DNA microarrays—comparison with Northern blot analyses. *Genomics*. 2001;71:34-39.
58. Bowtell DD. Options available—from start to finish—for obtaining expression data by microarray. *Nat Genet*. 1999;21:25-32.
59. Halgren RG, Fielden MR, Fong CJ, Zacharewski TR. Assessment of clone identity and sequence fidelity for 1189 IMAGE cDNA clones. *Nucleic Acids Res*. 2001;29:582-588.
60. Ichikawa JK, Norris A, Bangera MG, et al. Interaction of *Pseudomonas aeruginosa* with epithelial cells. *Proc Natl Acad Sci U S A*. 2000;97:9659-9664.
61. Wang K, Gan L, Jeffery E, et al. Monitoring gene expression profile changes in ovarian carcinomas using cDNA microarray. *Gene*. 1999;229:101-108.
62. Wong KK, Cheng RS, Mok SC. Identification of differentially expressed genes from ovarian cancer cells by MICROMAX cDNA microarray system. *Biotechniques*. 2001;30:670-675.
63. Grove DS. Quantitative real-time polymerase chain reaction for the core facility using TaqMan and the Perkin-Elmer/Applied Biosystems Division 7700 Sequence Detector. *J Biomol Techniques* [serial online]. 1999;10. Available at: <http://www.abrf.org/jbt/1999/March99/mar99index/html>. Accessibility verified October 9, 2001.
64. Raval P. Qualitative and quantitative determination of mRNA. *J Pharmacol Toxicol Methods*. 1994;32:125-127.
65. Jung R, Soondrum K, Neumaier M. Quantitative PCR. *Clin Chem Lab Med*. 2000;38:833-836.
66. Winzler EA, Shoemaker DD, Astromoff A, et al. Functional characterization of the *S cerevisiae* genome by gene deletion and parallel analysis. *Science*. 1999;285:901-906.
67. Garber K. Proteomics gears up. *Signals* [serial online]. 1999. Available at: <http://www.signalsmag.com/signalsmag.nsf/657b06742b5748e888256570005cba01/f8a34b7efde4eb6c8825681c000b8a96>. Accessibility verified October 16, 2001.
68. Shirota Y, Kaneko S, Honda M, et al. Identification of differentially expressed genes in hepatocellular carcinoma with cDNA microarrays. *Hepatology*. 2001;33:832-840.
69. Storz M, Zepter K, Kamarashev J, et al. Coexpression of CD40 and CD40 ligand in cutaneous T-cell lymphoma (mycosis fungoides). *Cancer Res*. 2001;61:452-454.
70. Dalton R, Abbott A. Can researchers find recipe for proteins and chips? *Nature*. 1999;402:718-719.

# 13

## Gene Expression Analysis Using Microarrays

Sophie E. Wildsmith and Fiona J. Spence

### 13.1 Introduction

Microarrays are of increasing interest to both industry and academia as tools for 'gene hunting' and also as quantitative methods for routine analysis of large numbers of genes. Techniques such as real-time polymerase chain reaction (RT-PCR) TaqMan™ and SybrMan™ are generally considered to be more accurate, robust, larger in dynamic range and less capital intensive, but for rapid, large-scale gene expression analysis using limited mRNA, microarrays and gene chips are preferred.

### 13.2 Microarray experiments

#### Platforms

Global gene expression platforms are now available in multiple formats, including cDNA arrays, oligonucleotides spotted onto slides or *in situ* synthesised oligonucleotide arrays manufactured using photolithography. Commercial sources for these include Stratagene (La Jolla, CA), Memorec (Köln, Germany) and BD BioSciences (Oxford, UK) for cDNA microarrays, Mergen Ltd (San Leandro, CA) for spotted oligomers and Affymetrix (Palo Alto, CA) for oligoarrays synthesised *in situ*. Purchasing from a supplier is more expensive than generating microarrays in-house, although the latter is beneficial in labour-intensive institutions or when proprietary gene information is utilised.



'Off-the-shelf' microarrays may also have the advantage of rigorous quality control and standardised protocols.

It is possible to produce oligonucleotide spotted arrays in-house, by designing oligonucleotide sequences that match genes of interest and then purchasing purified oligonucleotides to spot down on glass or other substrates. Alternatively there are new systems such as that available from CombiMatrix (Mukilteo, WA, USA) for computer-aided design and *in situ* synthesis of oligonucleotides. However, production of cDNA microarrays is currently the most affordable and popular method and is now well established. Numerous sources of information on cDNA microarray fabrication are available in the literature and on the internet (Bowtell, 1999; Cheung *et al.*, 1999; Wildsmith and Elcock, 2001 and <http://cmgm.stanford.edu/pbrown>). Thus, this chapter will focus on the implementation of experiments and analysis of data from cDNA microarrays. The experimental procedure differs slightly according to the number of fluorophores (or channels) and the type and manufacturer of the array. We have attempted to describe a generic process, indicating where possible the different options. Figure 13.1 demonstrates the procedure for a two-colour hybridisation.

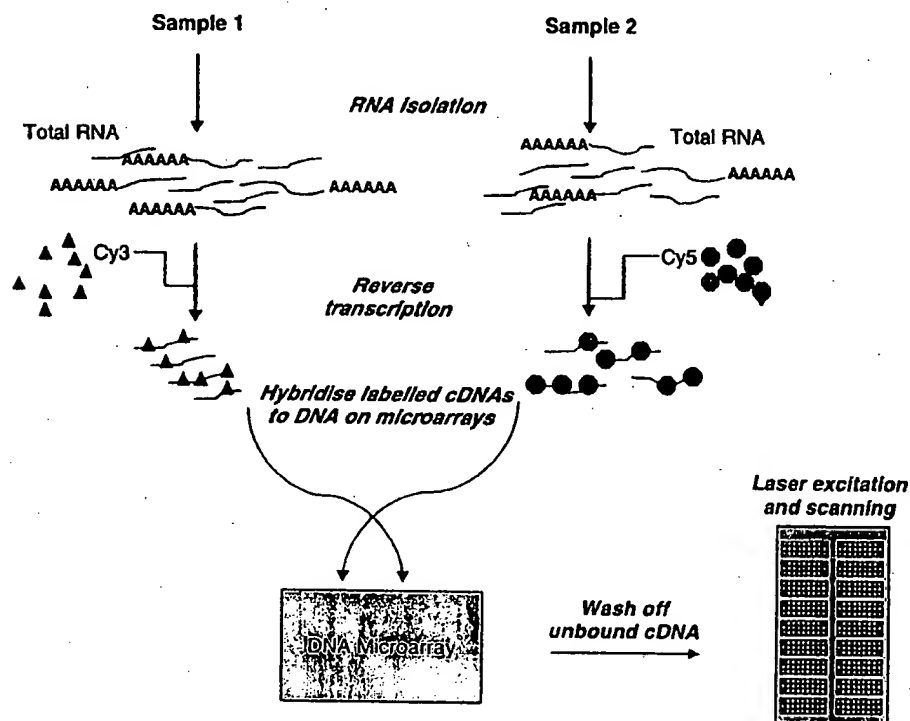


Figure 13.1 The microarray experimental process for two-colour hybridisations

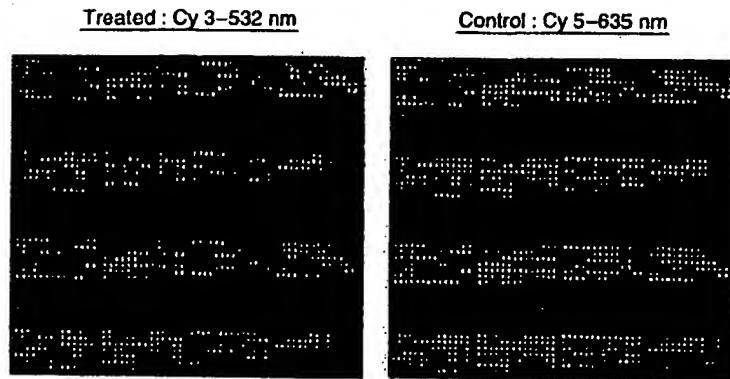


### RNA Extraction

First, RNA is extracted from the tissue or cells of interest. The quality of the RNA extracted is paramount to the overall success of the microarray experiment, as impurities in the sample can effect both the probe labelling efficiency and also stability of the fluorescent label (Hegde *et al.*, 2000). Snap-freezing of tissue in liquid nitrogen, immediately after harvesting, is used to preserve RNA integrity. Any further sectioning of the tissues should be carried out under RNase-free conditions (Fernandez *et al.*, 1997). Total RNA can be extracted using kits such as TRIzol® (Invitrogen, Paisley, Scotland) and Rneasy (Qiagen, GmbH, Hilden, Germany). Some researchers perform a further extraction of mRNA; this results in a purer starting material but has the disadvantage of lower yields. Affymetrix recommend between 5 and 40 µg of total RNA is required for their GeneChips™ and 10 µg or less is the required amount of starting material for cDNA microarrays (Hegde *et al.*, 2000).

### Sample Labelling

The mRNA is transcribed *in vitro*, with the concomitant inclusion of labelled nucleotides. The labels may be fluorescent or radioactive. In the case of dual channel/colour hybridisations, two samples will be labelled with dyes that fluoresce at different wavelengths, with different emission spectra. Example fluorophores, available coupled to nucleotides, are Cy3, Cy5, fluorescein and lissamine. Wildsmith *et al.* (2001) have demonstrated that AlexaFluor 546dUTP™ (Molecular Probes, Leiden, The Netherlands) gives a significantly higher signal than Cy3dCTP (Amersham Biosciences, Piscataway, NJ, USA). When performing two-colour hybridisations the control sample and 'test' sample are labelled with different fluorophores and the subsequent cDNA is then mixed together and hybridised simultaneously (Nuwaysir *et al.*, 1999). An advantage of simultaneously hybridising control and treated sample is that it obviates the need to control for differences in hybridisation conditions or between microarrays. A specific example of the huge impact this technique has had includes its use in the first published account of gene expression data of the entire genome of *Saccharomyces cerevisiae* (DeRisi *et al.*, 1997). In two-colour hybridisations, one would assume that the properties of the two fluorescent dyes being used are equivocal. In fact, for Cy5 and Cy3 this is not the case as Cy5 has been reported to give higher background fluorescence and also is more sensitive to photobleaching than Cy3 (Van Hal *et al.*, 2000). In addition, there is evidence from several independent sources that the combination of Cy3 and Cy5 dye labelling can affect data in certain genes. That is to say, when experiments are repeated and the dye combination for the two probes reversed, inconsistent results are obtained with certain genes (Taniguchi *et al.*, 2001). Despite these



**Figure 13.2** Two-colour fluorescent scan of human gene cDNA array. The probe mix consists of DNA from HepG2 control cells and cells treated with buthionine sulfoximine for 6 h. A colour version of this figure appears in the colour plate section

facts, two-colour hybridisations are widely accepted throughout the microarray community and an example of an image is shown in Figure 13.2.

### Hybridisation and Processing

After labelling, the cDNA is purified (to remove unincorporated nucleotides), mixed with a hybridisation buffer and then applied to a cDNA microarray slide. The sample and the slide are heated prior to hybridisation in order to separate double-stranded DNA. A coverslip is applied (Shalon *et al.*, 1996), or preferably a hybridisation chamber is used to avoid evaporation and enable an even hybridisation. The hybridisation and subsequent wash steps are carried out at a buffer stringency and temperature that enables hybridisation of complementary strands of DNA but reduces non-specific binding.

### Image Capture and Image Analysis

After hybridisation the microarray slides are scanned, using either a laser or a phosphorimager (depending on the type of label used). There are many different suppliers and models of fluorescence scanner, for example the ScanArray 5000 (Perkin Elmer Life Sciences, Zaventem, Belgium), GenePix 4000B (Axon GRI, Essex, UK) and the GeneArray® (Affymetrix, Santa Clara, CA). The choice of scanner is determined by sensitivity, resolution, flexible wavelength, file size generated, throughput and technical support available.

Images are analysed using software that measures the intensity of the signal from the hybridised spotted genes (spots), which provides a measurement of the amount of cDNA bound. Thus the initial concentration of messenger RNA is inferred. Early software packages 'drew' grids around the spots and

integrated across the whole area of the grid. This overcame problems associated with accurate location of the spots, which is problematic, especially if the spots on the printed arrays are poorly aligned. More recent versions of software 'draw' circles around the spots themselves and perform measurements within and outside of this boundary. For example, the background may be calculated from a region outside the spot boundary. The intensity of the signal from the spot may be calculated using median, mode or mean values of the pixels within the spot. Researchers differ in their preferences regarding using median or mean values (Hegde *et al.*, 2000) and this is likely to depend upon the protocols and software used.

Image analysis software commonly is supplied with scanners or can be bought from the same supplier. This has the advantage of being optimised for that specific type of microarray and the benefit of upgrades and technical support. Software for microarray analysis is available from BioDiscovery (<http://www.biodiscovery.com>), Imaging Research (<http://imagingresearch.com>), GenePix Pro (Amersham Biosciences, Piscataway, NJ), arraySCOUT 2.0 (<http://www.lionbioscience.com>), NIH ([http://www.nhgri.nih.gov/DIR/LCG/15K/HTML/img\\_analysis.html](http://www.nhgri.nih.gov/DIR/LCG/15K/HTML/img_analysis.html)), Stanford University (<http://rana.Stanford.EDU/software>) Media Cybernetics (Silver Spring, MD, USA) and TIGR (<http://www.tigr.org/softlab>). Important criteria for image analysis software include speed, ease of use, automation and the ability to distinguish artefact from real signal (Wildsmith and Elcock, 2001).

As the technology has evolved and more experience gained, it has become more and more apparent that the most significant issues facing microarray users are the processing of the vast quantities of data generated and deciding exactly what tools are the most appropriate for data analysis. Because of the enormity of this, we have dedicated a complete section to describing the current status of this area.

### 13.3 Data analysis

It is important to be cognisant of the fact that the practical laboratory aspects of using microarrays are only part of gene expression analysis. Many researchers generate vast volumes of data, without a clear understanding of how to manage and interpret them. Furthermore, the variability in microarray data confers additional problems for analysis. In some cases the purpose of the experiment will be a gene-hunting exercise, in which case a cursory indication of potential gene biomarkers is sufficient analysis. In other instances, such as pathway mapping and screening studies, it is paramount that results are statistically meaningful and valid. The next few sections detail some relatively simple analysis methods and recommendations for the benefit of researchers with minimal statistical

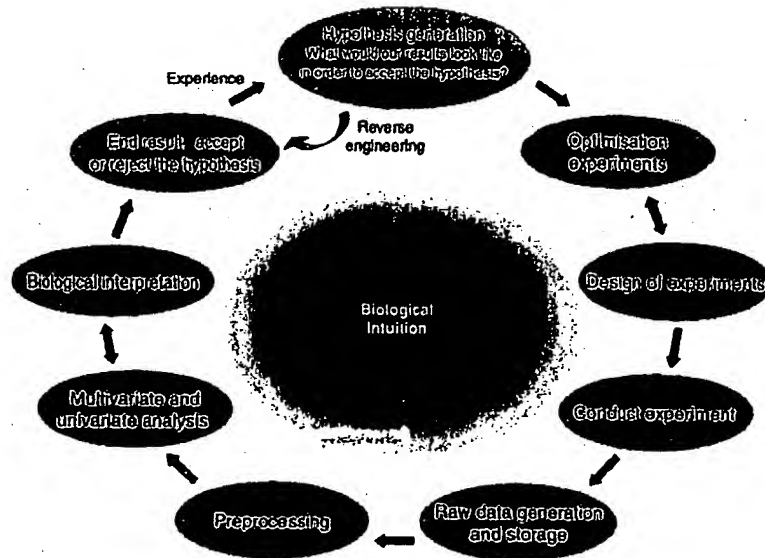


Figure 13.3 The ideal microarray experimental design and process

training. There are also suggestions for more advanced analysis for those who have the assistance of a statistician or specialist data analyst.

Most of the steps before, during and after performing a microarray experiment are optimally conducted with regard for statistics and data analysis. Careful planning before implementation facilitates the downstream analysis and interpretation of data. The following model summarises the entire microarray process with integration of the biological and data analysis components (Figure 13.3).

### Hypothesis Generation

Any study is conceived for the purpose of investigating or obtaining supporting evidence for a biological hypothesis. Giving time at this early stage to consider downstream implications will pay dividends later. It is helpful if, rather than simply stating the aims of the experiment, the researcher asks the question 'What results do I expect?' or 'what answer will validate/invalidate my hypothesis?'. This 'reverse-engineering' proves useful in focusing the project, assessing the feasibility of the work, providing early preparation for data management and analysis and, importantly, in managing expectations with regard to outcomes.

A good example of careful experimental planning is demonstrated by Golub *et al.* (1999) in the classification of acute leukaemias in order to distinguish between acute lymphoblastic leukaemia (ALL) and acute myeloid leukaemia (AML). Distinguishing between ALL and AML using conventional techniques is known to be a difficult task. The researchers maximised their probability of success by choosing an easier, more defined model (normal kidney vs. renal

cell carcinoma), on which to validate their analytical methods. In doing so they established that their techniques were suitable for classifying tissues according to disease and gained confidence in their approach before using the samples of real interest.

### Optimisation Experiments

Although microarrays are becoming increasingly accessible to all, using these tools requires experience and it is unlikely that successful experiments will be conducted immediately. It is usual that some time is given to optimising a system for any specific application, for example for a given tissue or cell type. Additionally, the requirements for a given system may warrant some modifications. The standard approach for a scientist to take is to vary one parameter, whilst keeping all others constant. This is time-consuming and does not take into account the interactions between different factors. Well-designed, multifactorial experiments (Box *et al.*, 1978), provide a faster route for optimisation, with a statistical measure of confidence. An example of this technique is in the optimisation of microarray experimental conditions for preparation of fluorescent probes from rat liver tissue (Wildsmith *et al.*, 2001). When a major source of variation is revealed this can be investigated further with a view to minimising it or providing sufficient replicates to account for it.

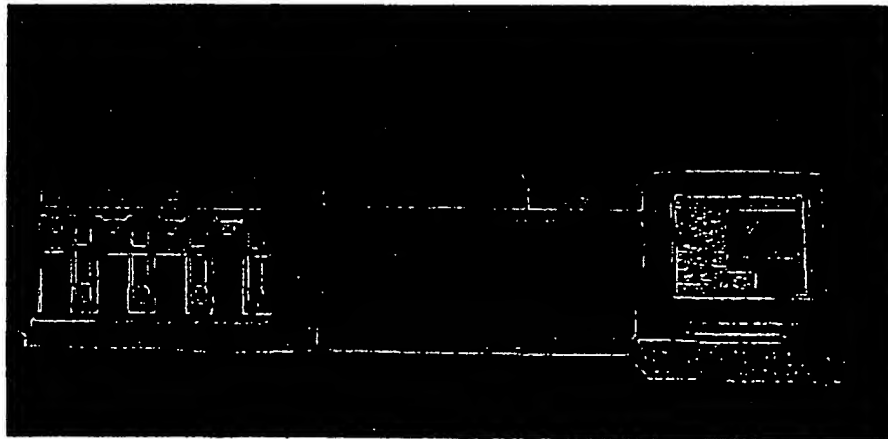
### Design of Experiments

Once confidence in the experimental procedure has been obtained the researcher is likely to have gained an insight into the reproducibility of the system. This assists in the design of the experiments, in particular in determining the minimal number of replicates necessary. Replication can be implemented at many stages – from biological samples through to microarray slides.

Owing to the enzyme-catalysed transcription reactions, a large amount of variation occurs during the probe-making stages in microarray experiments. Our work indicated that replicates should be made at this step and a minimum of six replicate probes are made for microarray experiments (Wildsmith *et al.*, 2001). These can be pooled or hybridised separately onto six microarray slides.

Lee *et al.* (2000) have examined the effect of the different location of cDNA spots on the glass slides and concluded that replicates are essential to provide meaningful data and to enable reliable inferences to be drawn.

With regard to commercially available gene chip systems, such as that available from Affymetrix (see Figure 13.4), the variation between chips, within a batch, is likely to be low due to stringent quality control and highly automated manufacture. The use of an automated wash station also reduces variability in intensities between chips. However, using a multi-step approach in the probe preparation and subsequent antibody binding steps may lead to variation between



**Figure 13.4** The GeneChip® Instrument System. From left to right, the hybridisation station, scanner and workstation. Image courtesy of Affymetrix. A colour version of this figure appears in the colour plate section

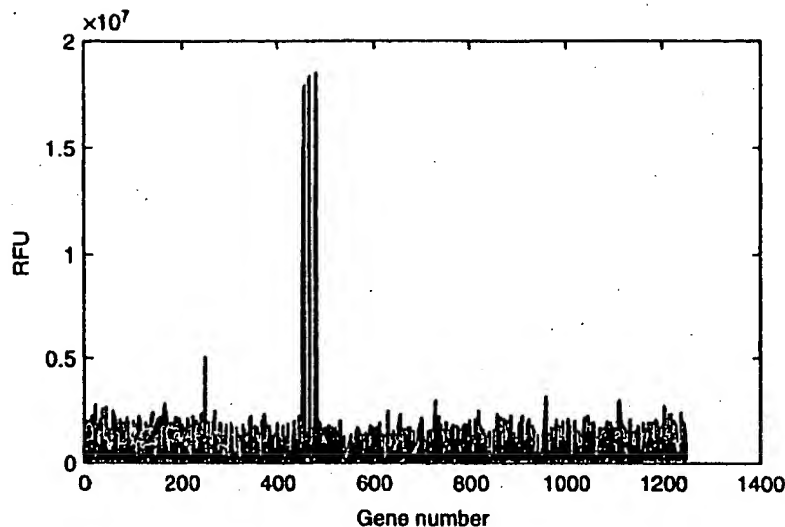
replicate samples prepared on different days. Pooling of reagents within an experiment, and analysing controls together with treated samples, will both reduce the variability within a given experiment.

### Conduct of Experiment

At this stage some attention may be required for verifying and validating processes. For example, checking that the imaging instruments give consistent results across the slide, on repeat use and from day to day. If two imagers are used it is important to verify that the results from both machines are comparable. Some laboratories read fluorescence of one channel and then adjust the laser intensity of the second channel in order to obtain comparable readings. This is a method of normalising for the difference in intensities of fluorophores. It is important to be aware that this approach has a number of drawbacks. The arbitrary value of the second laser intensity setting will vary from experiment to experiment; thus comparisons of this channel cannot be made across experiments. Also the response of the fluorophore may not be linear across the laser intensity settings and this can lead to additional errors.

Another area for investigation prior to running the study itself is the image analysis component. Depending on the software used, the image analysis package may process the data to some extent, for example automatic background subtraction. Full understanding of the software is required so that it is clear at what point the data are 'raw', and the extent of inherent, inseparable manipulation. Effort may be required to determine the optimum settings for any software parameters.

As data are generated it is important to be aware of the data integrity – for example ensuring that all data are collected, so that there are no missing data that

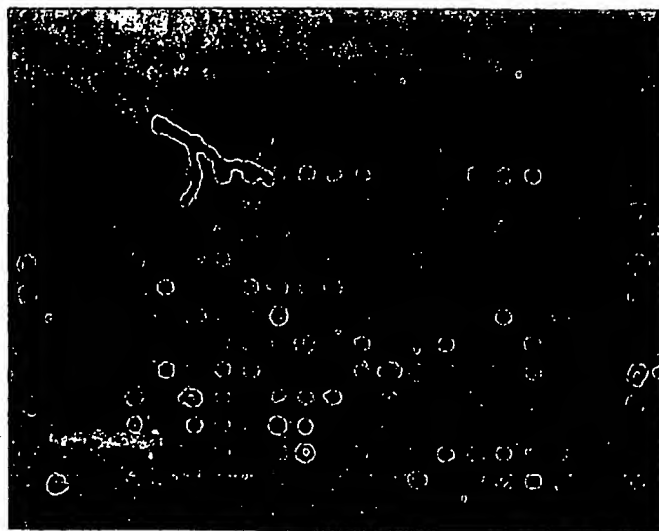


**Figure 13.5** The relative fluorescence units (RFUs) of 1248 genes on seven microarray slides that were hybridised with cDNA made from the liver of a rat treated with acetaminophen. Note the gene outliers at approximately gene number 480

can complicate analysis later. The researcher may be intuitively aware of any spurious results and should be alert for anything extraordinary that could indicate problems, for example hybridisation intensities appearing inconsistent from sample to sample. Data analysis at this point can be a rapid indicator of dubious results. For example, Figure 13.5 shows a plot of the fluorescent intensities of 1248 genes that were hybridised with probe derived from acetaminophen-treated rat liver tissue. The data appear consistent, with the exception of peaks in intensity on one slide at around gene 4800. Further investigation of the microarray revealed a large artefact that had been missed by the image analysis process (Figure 13.6).

### Raw Data Generation and Storage

One issue that arises when carrying out microarray analyses is how much data to store and in what form. For Good Laboratory Practice (GLP) purposes, often required in industry, storage of the raw data is necessary. This could be construed as the microarray image. Storing the image analysis results requires far less storage space and is easier to visualise, but it has the drawback that image analysis cannot be redone should superior software be available in the future. In reality, the methods used for microarrays are continually changing and the likelihood of revisiting old images on which the analysis has been performed, using outdated protocols is quite small.



**Figure 13.6** Portion of scanned image showing region where artefact occurred that caused very high signals, which were classed as outliers

### Pre-processing

A number of pre-processing steps are often used in microarray analysis. These include filtering, log transformation, normalisation and background subtraction. Filtering may be used before or after transformation in order to extract data from preferred regions of interest, or in order to remove outliers (see the above example relating to image analysis artefact). One example of filtering is the removal of individual gene replicates that lie outside a given number (for example 5) of standard deviations from the mean. Alternatively, data points that lie in the top/bottom few percentiles (e.g. 0.1%) of the data can be removed. This method of removing outliers is also called 'trimming'. It is acceptable if there is a large volume of data where only a small proportion of data is removed and if the same method is applied consistently across all data. Care must be taken in the way in which this is carried out in order not to delete genuine data. For example, if one gene is consistently high or low in expression across replicates, then it is unlikely to be an outlier.

Another method of detecting outliers is to plot all genes (see the section 'Conduct of experiment' above) or to perform PCA analysis (see the section 'Multivariate analysis' below) to detect replicate outliers. The use of a PCA plot to detect outliers is shown in Figure 13.7.

Log transformation of data is accepted universally because the fluorescence data that are generated from microarrays tend to be skewed towards lower values. There are scientifically valid reasons why ratios of raw expression values should not be used (Nadon and Shoemaker, 2002). When using two-colour



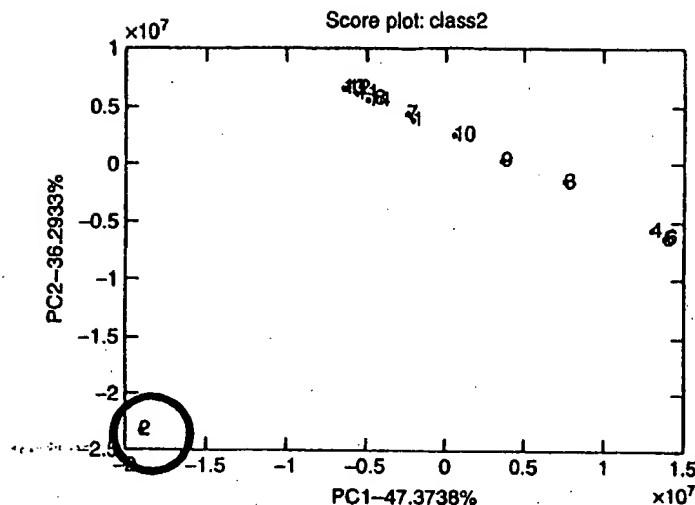


Figure 13.7 PCA plot of data used in Figure 13.4 showing one microarray (number 2) as an outlier

hybridisations it is common to express the ratio of treated to control as a logarithm in base 2 (Quackenbush, 2001). Thus genes up-regulated by a factor of 2 have a  $\log_2$  (ratio) of 1, and genes down-regulated by a factor of 2 have a  $\log_2$  (ratio) of  $-1$ .

Normalisation and background subtraction techniques are methods of data manipulation and their use is more subjective and often debated. The purpose of these techniques is to reduce the error (variability) that occurs between replicates and thus enable a comparison of data across samples.

The theory behind background subtraction is that during hybridisation there will be non-specific binding to the slide. This will effectively 'darken' the image and give falsely high readings of fluorescent intensity. Correcting for the non-specific hybridisation should reduce error due to background staining. Background subtraction often occurs automatically in microarray image analysis packages. The software may circle the spot of interest and use the region beyond the periphery as the measurement of background. In cases of uneven hybridisation this method enables locally high background to be subtracted on a regional basis. One criticism of this approach is that the slide surface beyond the periphery is not similar, in chemical terms, to that where the nucleic acid has been deposited, and therefore cannot act as a real control for non-specific binding. A more accurate measurement of non-specific binding can be gained from using a region where spotting chemicals have been deposited, but no target is present. This concept is the basis of a method using local 'blank spots' (Wu *et al.*, 2001).

A number of methods exist for normalisation of data. These include normalising to total signal or to a 'known' spot or gene, standardisation, or proprietary

methods. Normalising to total signal is the simplest approach, whereby the gene intensity is expressed as a percentage or proportion of the signal intensity for the entire array. This method works best when the total intensities for the microarrays are similar and the number of changes is small compared with the number of genes. However, we often find, when using arrays of around 1000 genes, that pathological disease can up-regulate a large number of genes simultaneously. In this case, when total signal normalisation is applied, highly up-regulated genes will appear less up-regulated and genes that do not change from the control will appear down-regulated.

Normalisation to a control value is a more popular technique. A control value can be obtained from using a gene known to remain constant under the conditions of the experiment. DeRisi *et al.* (1997) used a panel of 90 housekeeping genes for normalisation, but found considerable variation in their gene expression. Unfortunately it is very difficult to know with certainty that a gene will not change and there is evidence to suggest that so-called 'housekeeping genes' are variable (Savonet *et al.*, 1997). Other control genes can be derived from an alternative species; these should not be expected to hybridise. We have used yeast and Arabidopsis genes as negative controls for hybridisations of rat tissues. No orthologs were known to the genes selected; however, in most cases non-specific binding occurred.

If two or more microarray replicates appear to be different, but they are expected to be the same, then they can be standardised. An example of this might occur if the total intensity of one microarray is greater than another, but the genes are proportionally equally up- or down-regulated. If the microarray sample spot data are assumed to be drawn from a normal distribution, then the 'z-transform' can be used. This requires that the mean and the standard deviation of the intensity values for each microarray are determined. The mean is subtracted from each individual gene value and the remainder is divided by the standard deviation. The intensity values from each microarray will then have the same mean and standard deviation. This has the advantage of facilitating comparison of microarrays with different dynamic ranges as well as total intensities. If the data are not normally distributed, then alternative non-parametric methods can be used, such as normalising to the median.

### Univariate Analysis

Univariate methods of analysis involve examining one variable, or gene, at a time. This can be a very laborious task when examining a large volume of data, yet it is the preferred method of biologists. The simplest technique is to compare control and treated values and express the result as a 'fold-change' ratio. Typically, when examining small volumes of data, fold changes greater than 2 and less than 0.5 are considered meaningful (Quackenbush, 2001). This cut-off is essentially arbitrary and has the distinct drawback that microarray data

are not homoscedastic; that is, there is more variation about the mean at low values than there is at high values (Draghici, 2002).

A second method for finding up- or down-regulated genes uses the standard deviation (SD) of the replicate gene data. Thus if changes greater than, say, 2 SD from the log mean ratio are considerably greater than changes associated with 'noise', then they are considered significant. This technique means that when looking at a large number of genes that are normally distributed there will be up- and down-regulated genes, regardless of whether there are (biological) changes (Draghici, 2002).

Rather than using arbitrary cut-off values it is far more meaningful to express the fold-change in terms of either confidence intervals, or a '*p*-value' (that is, the probability of the value occurring by chance). Thus, a fold-change of 1.1 may be associated with a *p*-value of 0.001 and thus the probability that the gene is *not* up-regulated is 1 in 1000. Naturally, such small fold-changes may then be queried in terms of biological significance. One must then ask the question: Are large fold-changes more important (biologically) than small ones? Simple calculation of *p*-values for two data sets can be obtained using *t*-test functions in standard spreadsheet software. A number of replicates are necessary for this approach, and the data must be normally distributed. We have recently developed a method for calculating *p*-values for fold-changes that is not influenced by the distribution of the data or outliers and applied it to TaqMan™ and microarray data. Other complex and computationally intensive methods for calculating *p*-values are described in Draghici (2002) and Nadon and Shoemaker (2002).

### Multivariate Analysis

Multivariate analysis of gene expression data is becoming increasingly popular in the microarray community and in other biological domains where large volumes of data are generated. Multivariate analysis methods include principal component analysis (PCA), factor analysis, multivariate analysis of variance (MANOVA) and cluster analysis. Currently, cluster analysis is the most widely-used method in the microarray community but PCA is growing in popularity (Crescenzi and Giuliani, 2001; Konu *et al.*, 2001).

Quackenbush (2001) provides a good review of clustering tools that is rather unique in the regard that different clustering algorithms and linkage methods are presented. Clustering methods are unsupervised, and they are powerful tools for gaining insight into huge data sets. They enable the data to be partitioned in order to facilitate interpretation; however, they do suffer from subjectivity. This is because the user selects various parameters, such as the algorithm used, linkage type, distance metric and, sometimes, cluster size. Whatever the data, clusters will always be identified, thus there is also a tendency to over-interpret the data – trying to attach meaning to clusters that may have no biological relevance.

Software available for cluster analysis includes Cluster, the output of which is viewed in Treeview; both available from <http://rana.stanford.edu/software>. This tool is particularly useful for clustering genes to identify genes that are co-regulated.

The PCA is a visualisation tool that enables complex, high-dimensional data to be represented in two or three dimensions. It facilitates identification of groups of similar data, thus enabling inferences to be made about the samples.

An example is shown in Figure 13.8. The figure shows the gene data for one sample (control rat liver) that was hybridised to seven microarrays according to the method used in Wildsmith *et al.* (2001). Each microarray contained two replicate gene sets; thus there were 14 replicate gene sets in total. The gene sets comprised 1248 genes (with controls). All the data ( $14 \times 1248$  data points) was input into the analysis and the PCA plot displays the 14 replicates individually. The two axes are principal components 1 and 2. Principal component 1 (PC1) accounts for 65.5% of the variation in the data, whereas PC2 represents only 13%. This means that the model accounts for 78.5% of variation in the data.

The first principal component (PC1) accounts for as much as possible of the variation in the original data and subsequent components (e.g. PC2) are of decreasing importance. Thus, samples 8 and 9 are very different from samples 1 and 2. In terms of interpreting the PCA plot, it is immediately clear that there are three or four distinct clusters of data. These are marked by circles. Datapoints tend to cluster in pairs; for example replicates 1 and 2, 3 and 4, 5 and 6, etc. These are the duplicate gene sets on the same microarray. This indicates that the variation within the microarray is lower than the variation between

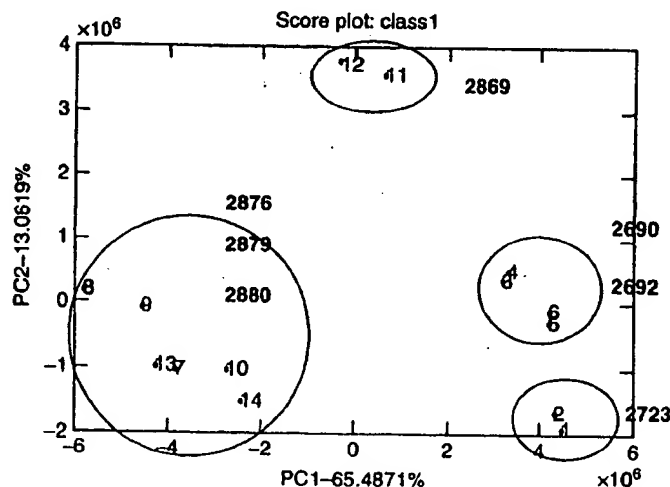


Figure 13.8 PCA plot of the microarray results from seven slides (2690, 2692, 2723, 2869, 2876, 2879, 2880), each with two replicate spot sets (labelled 1–14), after hybridisation with control rat liver. See text for explanation

replicate microarrays. However, given that the same sample is applied to all the microarrays, we must ask why we get further separation of the replicates. The answer lies in the associated data. The four-digit numbers associated with the clusters are the microarray slide numbers that indicate when they were printed. Numbers that are more similar seem to be more closely related, and thus we can hypothesise that there were some differences between the slides, such as differences in slide backgrounds or changes during the printing process or change-over of batches, between slides in the 2600–2700 region and the 2800s.

The PCA provides a clearer overview of the data than does cluster analysis. It is a rapid method for gaining an insight into the results, in particular where biological meaning can be attached to the components (Crescenzi *et al.*, 2001). There are a number of packages for multivariate data analysis, including SIMCA-P (Umetrics, AB, Umea, Sweden) and The Unscrambler (Camo ASA, Norway), both of which are useful for PCA.

Other tools for data visualisation include software packages such as Spotfire.net (Spotfire Inc., Cambridge, MA, USA) and GeneSpring (Silicon Graphics, San Carlos, CA, USA). Spotfire is particularly useful for visualisation of multidimensional data and for visualisation of temporal data. It is possible to use these tools to identify genes that are co-ordinately expressed over time.

### Biological Interpretation

After developing a sound experimental strategy, ensuring that the results are statistically valid, and after analysis of the data, it is down to the biologist to assemble the pieces of information that have been obtained. This intertwined information may include unexpected results that are contradictory to intuition or to published literature. One way to untangle the data is to map the relevant genes onto existing pathways and known functions. The Kyoto Encyclopedia of Genes and Genomes (KEGG), available at <http://www.genome.ad.jp/kegg/>, is a useful source of information, especially where the gene products are enzymes. It enables visualisation of the position of up- or down-regulated genes in metabolic pathways.

The gene expression data obtained may differ from protein expression data, or information on gene product activity or location. When initiating a study it is useful to consider additional endpoints that can assist in the interpretation of the data. For *in vitro* studies, these might include cytotoxicity endpoints, metabolites, key signalling molecules or perhaps protein expression. Waring *et al.* (2001) used tetrazolium dye reduction (MTT) as a measure of hepatocyte cell viability for their studies of gene expression in response to hepatotoxic insult. For *in vivo* studies, expression information on the tissue of interest could be supported by pathology, histology and blood chemistry measurements. Gene expression results could be confirmed by *in situ* hybridisations or protein activity assays.

### 13.4 Recent examples of microarray applications

One area of rapid progress using microarray technology is the increased understanding of cancer. Molecular pathologists are subgrouping cancers of tissues such as blood, skin and breast, based on differential gene expression patterns. For example, within a small group of breast cancer tissue samples, Perou *et al.* (2000) distinguished two broad subgroups representing those expressing or alternatively lacking expression of the oestrogen receptor- $\alpha$  gene. The work was not conclusive, but never has progress in this field been so rapid when compared with the previous methods of gene identification.

Another example of the impact of this technology is in the identification of two biomarkers for prostate cancer, namely hepsin and PIM1 (Dhanasekaran *et al.*, 2001).

Microarray technology has also accelerated the understanding of the molecular events surrounding pulmonary fibrosis. Specifically, two distinct clusters of genes associated with inflammation and fibrosis have been identified in a disease where, for years, the pathogenesis and treatment have remained unknown (Katsuma *et al.*, 2001).

### 13.5 Conclusions

Important factors in gene expression experiments include sensitivity, precision and reproducibility in the measurement of specific mRNA sequences (Schmittgen *et al.*, 2000). These quality metrics can be maximised by using, or fabricating, high-quality microarrays, and by optimising each step of the microarray process. From conception to conclusion it is important to bear in mind the original hypothesis.

Having considered the complexity of the microarray experiment, the value obtained from a meticulously designed experiment should not be underestimated. As the number of high-quality gene expression studies increases, we hope that the literature will contain increasingly detailed information that will help interpret complex gene expression changes, and thus elucidate the mechanisms of disease.

### 13.6 Acknowledgements

The authors express their thanks to W. Wu for Figures 13.7 and 13.8, and to W. Wu, S. Roberts and P. Lane for useful comments and discussions.

### 13.7 References

- Bowtell DL (1999) Options available – from start to finish – for obtaining expression data by microarray. *Nature Genet.* 21: 25–32.

- Box GEP, Hunter WG and Hunter JS (1978) Chapter 7, in: *Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building*. John Wiley and Sons, New York.
- Cheung VG, Morley M, Aguilar F *et al.* (1999) Making and reading microarrays. *Nature Genet.* 21: 15–19.
- Crescenzi M and Giuliani A (2001) The main biological determinants of tumor line taxonomy elucidated by a principal component analysis of microarray data. *FEBS Lett.* 507: 114–18.
- DeRisi JL, Iyer VR and Brown PO (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278: 680–6.
- Dhanasekaran SM, Barrette TR, Ghosh D *et al.* (2001) Delineation of prognostic biomarkers in prostate cancer. *Nature* 412: 822–4.
- Draghici S (2002) Statistical intelligence: effective analysis of high-density microarray data. *Drug Discov. Today*, in press.
- Fernandez PM, Pluta LJ, Fransson-Steen R *et al.* (1997) Reverse transcription-polymerase chain reaction-based methodology to quantify differential gene expression directly from microdissected regions of frozen tissue sections. *Mol. Carcinog.* 20: 317–26.
- Golub TR, Slonim DK, Tamayo P *et al.* (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286: 531–7.
- Hegde P, Qi R, Abernathy K *et al.* (2000) A concise guide to cDNA microarray analysis. *Biotechniques* 29(3): 548–62.
- Katsuma S, Nishi K, Tanigawara K *et al.* (2001) Molecular monitoring of bleomycin-induced pulmonary fibrosis by cDNA microarray-based gene expression profiling. *Biochem. Biophys. Res. Commun.* 288: 747–51.
- Konu O, Kane JK, Barrett T *et al.* (2001) Region-specific transcriptional response to chronic nicotine in rat brain. *Brain Res.* 909: 194–203.
- Lee TML, Kuo FC, Whitmore GA *et al.* (2000) Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridisations. *Proc. Natl. Acad. Sci. USA* 97(18): 9834–9.
- Nadon R and Shoemaker J (2002) Statistical issues with microarrays: processing and analysis. *Trends Genet.* 18(5): 265–71.
- Nuwaysir EF, Bittner M, Trent J *et al.* (1999) Microarrays and toxicology: the advent of toxicogenomics. *Mol. Carcinog.* 24: 153–9.
- Perou CM, Sorlie T, Eisen MB *et al.* (2000) Molecular portraits of human breast tumours. *Nature* 406: 747–52.
- Quackenbush J (2001) Computational analysis of microarray data. *Nature Rev. Genet.* 2: 418–27.
- Savonet V, Maenhaut C, Miot F *et al.* (1997) Pitfalls in the use of several 'housekeeping' genes as standards for quantitation of mRNA: the example of thyroid cells. *Anal. Biochem.* 247: 165–7.
- Schmittgen TD, Zakrajsek BA, Mills AG *et al.* (2000) Quantitative reverse transcription-polymerase chain reaction to study mRNA decay: comparison of endpoint and real-time methods. *Anal. Biochem.* 285: 194–204.
- Shalon D, Smith SJ and Brown PO (1996) A DNA microarray system for analysing complex DNA samples using two-colour fluorescent probe hybridisation. *Genome Res.* 6: 639–45.
- Taniguchi M, Miura K, Iwao H *et al.* (2001) Quantitative assessment of DNA microarrays – comparison with northern blot analyses. *Genomics* 71: 34–9.
- Van Hal NLW, Vorst O, Van Houwelingen MML *et al.* (2000) The application of DNA microarrays in gene expression analysis. *J. Biotech.* 78: 271–80.
- Waring JF, Ciurlionis R, Jolly RA *et al.* (2001) Microarray analysis of hepatotoxins *in vitro* reveals a correlation between gene expression profiles and mechanisms of toxicity. *Toxicol. Lett.* 120: 359–68.
- Wildsmith SE and Elcock FJ (2001) Microarrays under the microscope. *J. Mol. Pathol.* 54: 8–16.

- Wildsmith SE, Archer GEB, Winkley AJ *et al.* (2001) Maximisation of signal associated with cDNA microarrays. *BioTechniques* 30: 202–8.
- Wu W, Wildsmith SE, Winkley AJ *et al.* (2001) Chemometric strategies for normalisation of microarray data. *Anal. Chim. Acta.* 446: 449–64.

### 13.8 Further reading

- Chatfield C and Collins AJ (1980) *Introduction to Multivariate Analysis*. Chapman & Hall, London.
- Heck DE, Roy A and Laskin JD (2001) *Nucleic Acid Microarray Technology for Toxicology: Promise and Practicalities. Biological Reactive Intermediates VI*. Kluwer Academic Press.
- Schena M (ed.) (2001) *Microarray Biochip Technology*. Eaton Publishing.

### 13.9 Useful websites

Protocols:

<http://cmgm.stanford.edu/pbrown>  
<http://cmgm.stanford.edu/pbrown/protocols.html>

Image analysis software:

<http://www.genome.ad.jp/kegg/>  
<http://rsb.info.nih.gov/nih-image>  
<http://rana.stanford.edu/software>  
[http://www.nhgri.nih.gov/DIR/LCG/15K/HTML/img\\_analysis.html](http://www.nhgri.nih.gov/DIR/LCG/15K/HTML/img_analysis.html)  
<http://rana.Stanford.EDU/software>  
<http://www.tigr.org/softlab>



## ORIGINAL ARTICLE

# Gene expression signatures and biomarkers of noninvasive and invasive breast cancer cells: comprehensive profiles by representational difference analysis, microarrays and proteomics

GM Nagaraja<sup>1</sup>, M Othman<sup>2</sup>, BP Fox<sup>1</sup>, R Alsaber<sup>1</sup>, CM Pellegrino<sup>3</sup>, Y Zeng<sup>2</sup>, R Khanna<sup>2</sup>, P Tamburini<sup>3</sup>, A Swaroop<sup>2</sup> and RP Kandpal<sup>1</sup>

<sup>1</sup>Department of Biological Sciences, Fordham University, Bronx, NY, USA; <sup>2</sup>Department of Ophthalmology and Visual Sciences, University of Michigan, Ann Arbor, MI, USA and <sup>3</sup>Bayer Corporation, West Haven, CT, USA

We have characterized comprehensive transcript and proteomic profiles of cell lines corresponding to normal breast (MCF10A), noninvasive breast cancer (MCF7) and invasive breast cancer (MDA-MB-231). The transcript profiles were first analysed by a modified protocol for representational difference analysis (RDA) of cDNAs between MCF7 and MDA-MB-231 cells. The majority of genes identified by RDA showed nearly complete concordance with microarray results, and also led to the identification of some differentially expressed genes such as lysyl oxidase, copper transporter ATP7A, EphB6, RUNX2 and a variant of RUNX2. The altered transcripts identified by microarray analysis were involved in cell–cell or cell–matrix interaction, Rho signaling, calcium homeostasis and copper-binding/sensitive activities. A set of nine genes that included GPCR11, cadherin 11, annexin A1, vimentin, lactate dehydrogenase B (upregulated in MDA-MB-231) and GREB1, S100A8, amyloid  $\beta$  precursor protein, claudin 3 and cadherin 1 (downregulated in MDA-MB-231) were sufficient to distinguish MDA-MB-231 from MCF7 cells. The downregulation of a set of transcripts for proteins involved in cell–cell interaction indicated these transcripts as potential markers for invasiveness that can be detected by methylation-specific PCR. The proteomic profiles indicated altered abundance of fewer proteins as compared to transcript profiles. Antisense knockdown of selected transcripts led to inhibition of cell proliferation that was accompanied by altered proteomic profiles. The proteomic profiles of antisense transfectants suggest the involvement of peptidyl-prolyl isomerase, Raf kinase inhibitor and 80 kDa protein kinase C substrate in mediating the inhibition of cell proliferation.

*Oncogene* (2006) 25, 2328–2338. doi:10.1038/sj.onc.1209265; published online 28 November 2005

**Keywords:** representational difference analysis; microarrays; proteomics; breast carcinoma; biomarkers; copper homeostasis

## Introduction

The transformation of a normal cell into a cancer cell has been correlated to altered expression of a variety of genes (Perou *et al.*, 2000; Becker *et al.*, 2005). The expression of some of these genes is a direct result of sequence mutation, whereas other changes occur due to alterations in gene products that participate in specific pathways. The changes in gene expression have been routinely characterized by classical subtraction hybridization and differential display approaches (Cerosaletti *et al.*, 1995; Alpan *et al.*, 1996). With the availability of the human genome sequence and sequences for a number of other model organisms, traditional methods have largely been replaced by gene microarrays (Khan *et al.*, 2001). These analyses have been used to characterize the molecular basis of a variety of diseases including cancer. A comprehensive analysis of a large number of cancer cell lines allowed clustering of genes into groups based on their expression patterns in phenotypically related cell lines (Khan *et al.*, 2001; Dan *et al.*, 2002; Rosenwald *et al.*, 2002; van't Veer *et al.*, 2002). The results of profiling experiments indicated expression of specific gene clusters in cell lines that have the same origin or have arisen from the same organ (Ross *et al.*, 2000). A complementary approach that has been used in limited ways is proteomics. Proteomics scores for changes in different proteins and peptides in cells with characteristic phenotypic differences. However, a comparative analysis of transcripts and proteins to establish a relationship between transcript changes and protein levels has not yet become routine.

Although expression profiling of tumor tissue and its comparison with normal tissue, in principle, is most appropriate to obtain the genetic signatures of a tumor type, such comparisons have not been free of attendant

Correspondence: Dr RP Kandpal, Department of Biological Sciences, Larkin Hall, Room 250, 441 E Fordham Road, Fordham University, Bronx, NY 10458, USA.

E-mail: kandpal@fordham.edu

Received 29 September 2005; revised 14 October 2005; accepted 14 October 2005; published online 28 November 2005

complications. These complications arise due to heterogeneity of tumor specimens wherein any cell type-specific changes are likely to be masked by other cell types that constitute the tumor specimen. For this reason, well-characterized cell lines established from tumor tissue may prove more informative and have been considered useful by cancer researchers. Comparing gene profiles between cell lines has the potential to reveal genes that could be causative for the phenotype and other genes that can serve as tumor biomarkers.

Our investigations are aimed at designating a subset of transcripts that could distinguish a normal breast cell from a breast cancer cell and help to predict tumorigenic or metastatic potential of a transformed cell. We describe here transcript and proteomic profiles of a normal breast cell line, a tumorigenic but noninvasive breast carcinoma cell line and an invasive breast carcinoma cell line, and summarize them as a set of candidate biomarkers or targets for therapeutic intervention. The comparison of transcript profiles with proteomic profiles demonstrated that altered proteins were not always represented in the microarray designated profiles and *vice versa*. Furthermore, we have targeted five transcripts that were upregulated in MCF7 cells for investigating their role in cell proliferation pathways. The proteomic profiles have revealed that inhibition of cell proliferation by antisense knockdown was mediated by a specific set of proteins.

## Results

### Representational difference analysis

As described in the Materials and methods section, RDA was performed by using cDNAs from MCF7 and MDA-MB-231 as tester/driver or driver/tester combination. The difference product in the first case represents the genes that are either upregulated in or specific to MCF7. On the other hand, the difference product of MDA-MB-231 (tester) and MCF7 (driver) hybridization resulted in the isolation of cDNAs that are either upregulated in or specific to MDA-MB-231. The initial linkers used in this protocol had internal *Bgl*III sites. One strand of the linker was used to amplify both the tester and driver cDNAs after linkers had been ligated to cDNAs. After removal of linkers from amplified cDNAs by digestion with *Bgl*III, a dephosphorylated *Bgl*III adaptor was ligated to tester DNA. The *Bgl*III adaptor had an internal *Eco*RI site. The difference product was digested with *Eco*RI and cloned in pBlue-Script vector. The cloning efficiency of the difference product was very low ( $5 \times 10^4$  c.f.u./ $\mu$ g of DNA). The low efficiency of cloning is attributed to a substantial fraction of amplified DNA product that is refractory to restriction digestion. The sequencing of a set of 100 clones each from the difference libraries revealed 50 different kinds of clones. The majority of these sequences were short fragments and represented either 3' regions or internal fragments of transcripts. A summary of these clones is presented in Table 1. The

involvement of the majority of these cDNAs is well characterized either in tumorigenesis or in metastasis. The phenotypic characteristics of MCF-7 and MDA-MB-231 ideally match with the biological significance of these genes. The alterations in transcripts for Rho signaling proteins,  $\text{Ca}^{2+}$  binding/requiring proteins, tight junctions/anchoring junctions/gap junctions, copper binding or sensitive proteins, and RUNX2 are particularly noteworthy.

The differential expression of a representative number of RDA clones was validated by semiquantitative PCR. As shown in Figure 1, these transcripts were either specific to or upregulated in the cell line that was used as a tester. Such analyses demonstrated that more than 90% of the clones were differentially expressed. The abundance of transcripts and the results of RT-PCR were also confirmed by Northern blotting (Figure 2). The pattern of hybridization clearly indicates that all these transcripts showed differential expression in MCF7 and MDA-MB-231 cells that were used as driver/tester combinations for the RDA.

### Gene microarrays

After obtaining preliminary molecular signatures of MCF7 and MDA-MB-231 cells by RDA, we used Affymetrix gene microarrays to expand the above analysis to identify a comprehensive set of transcripts that is deregulated in invasive breast carcinoma cells. The comparisons of cell lines on the basis of transcripts that are either present or absent as shown in Figure 3 revealed that a set of 123 genes distinguishes MDA-MB-231 cells from MCF7 and MCF10A. These genes can be classified by their involvement in functional classes such as transcription, signal transduction, cell adhesion, cell cycle, metabolism, transport, response genes and development (Figure 4). The majority of these genes participated in the process of signal transduction followed by transcription, cell adhesion and metabolism, respectively. A few transcripts in these classes were tested by real-time RT-PCR to confirm their altered abundance. The selected transcripts showed changes ranging between two- and 10-fold, 11- and 20-fold and greater than 20-fold, and were in close agreement with the results of microarray analysis. The qualitative pattern of change observed in microarrays analysis was readily reproduced by real-time or semiquantitative RT-PCR for all transcripts tested.

The number of altered transcripts was over 1000 based on a change of twofold or greater, and a majority of these genes show changes varying between two- and fourfold (Figure 5). Interestingly, with all comparisons combined, there were 21 genes downregulated more than 50-fold and 55 genes that were upregulated more than 18-fold when specific cell line pairs were compared (Figure 5). The transcripts that represent the extremes of upregulated and downregulated scale can allow distinction between MCF7 and MDA-MB-231 cells. These transcripts include GPCR11, cadherin 11, annexin A1, vimentin, lactate dehydrogenase B (upregulated in MDA-MB-231) and GREB1, S100A8, amyloid

**Table 1** Differentially expressed transcripts identified by RDA

<i>Upregulated in MDA-MB-231</i>	<i>Downregulated in MDA-MB-231</i>
<i>Extracellular matrix/matrix-crosslinking proteins</i> LOX Laminin $\beta$ 1 Collagen VI- $\alpha$ 1	<i>Calcium-binding proteins</i> Calgranulin B
<i>Calcium-binding proteins</i> Reticulocalbin 1 S100A8 Cullin 5	<i>Transcription factors/promoter-binding proteins:</i> Chromosome 4 ORF Estrogen receptor 1 RUNX2 variant (exon 8 deleted)
<i>Transcription factors/promoter-binding proteins</i> RUNX2 c-Jun Fra-1	<i>Cell-cell adhesion/cell-surface receptor proteins</i> Claudin 3 Amyloid $\beta$ precursor protein Triose phosphate isomerase Plakoglobin Cdh 1 Cdh 3 Annexin A9 RAR- $\alpha$ Connexin 31
<i>Cell-cell adhesion/cell-surface receptor proteins:</i> Cdh11 CYR61 MHC class II antigen $\gamma$ chain Protease-activated receptor-1 Protease-activated receptor-4	<i>ATPase/GTPase and signal transduction proteins</i> RhoD RhoB TGF- $\beta$ R1
<i>ATPase/GTPase and signal transduction proteins</i> ATP7a Caveolin 2 AXL receptor tyrosine kinase Rho GEF 3 Rho/Rac GEF 18 P21-Rac2	<i>Stress-response proteins</i> Protein kinase H11
<i>Metalloproteases and MMP inhibitor proteins</i> TIMP-2 MT1-MMP	<i>Cytoskeletal component and binding proteins</i> Keratin 18 Tubulin $\delta$ 1 Microtubule-associated protein $\tau$
<i>Stress-response proteins</i> Dual specificity phosphatase (DUSP)	<i>Cell-cycle regulation and growth/differentiation/apoptosis proteins</i> S100A13 S100C Aurora kinase AIK2 Nucleosidediphosphate kinase
<i>Cytoskeletal component and binding proteins</i> Moesin Vimentin Filamin B	<i>Secreted proteins and growth factors</i> Trefol factor 3 (TFF3) Trefol factor 1 (TFF1) Four and a half LIM domain1 Solute carrier family 16 SLC16A6
<i>Cell-cycle regulation and growth/differentiation/apoptosis proteins</i> Cyclin B1 Cyclin E Cyclin A2 Bcl2-like 1 protein	<i>DNA replication</i> DNA replication complex GINS-PSF2
<i>Secreted proteins and growth factors</i> Milk fat globule protein TGF- $\alpha$ SMAD-specific E3 ub ligase 2	<i>Miscellaneous</i> Serine protease inhibitor type 1 (SPINT1) Human homolog of <i>Xenopus</i> protein XAG Hypothetical protein FLJ22222 Hypothetical protein 20171 Hypothetical protein MGC3265
<i>Miscellaneous</i> Prion protein	

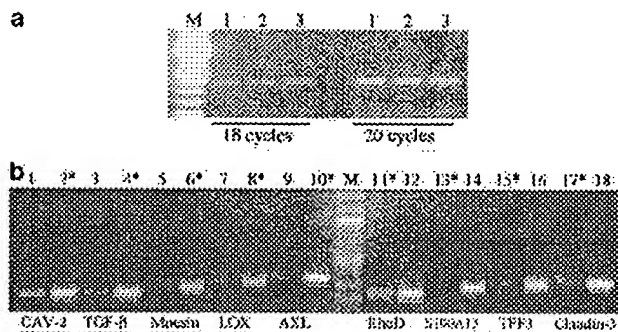
$\beta$  precursor protein, claudin 3 and cadherin 1 (down-regulated in MDA-MB-231). The distinction between MCF7 and MCF10A may be made based on keratin 19, serine protease, amyloid  $\beta$  precursor, neuropeptide Y receptor Y1 (upregulated in MCF7) and caldesmon, annexin A1, epithelial membrane protein 1, S100A2,

keratin 15 (downregulated in MCF7). Likewise, MDA-MB-231 cells differ from MCF10A in vimentin, epithelial membrane protein 3, cadherin 11, GPCR 116, collagen type XIII  $\alpha$  1, Bcl2-associated athanogene 2 (upregulated in MDA-MB-231) and keratin 15, cystatin A, cadherin 1, CD24, calcium-activated chlor-

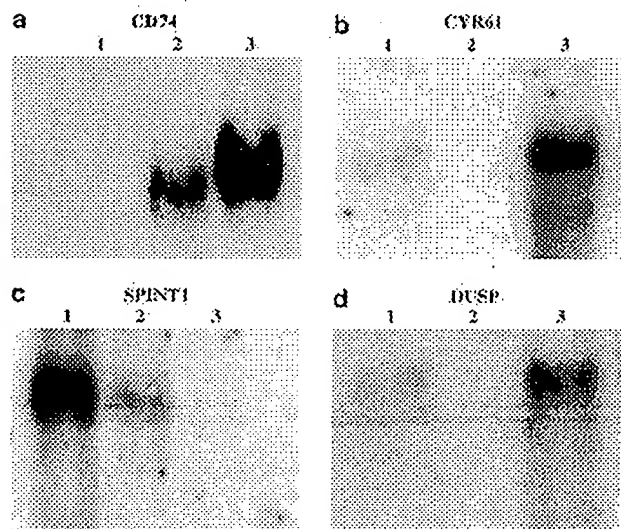
ide channel, S100P, GPCR 87 (downregulated in MDA-MB-231). Thus, a small subset of transcripts may serve as an accurate signature of these cell lines. Several of these gene products have been shown to participate in tumorigenesis and invasiveness of breast carcinoma cell lines.

The invasiveness phenotype of MDA-MB-231 cells specifically relates to changes in the following functional classes: (a) cell adhesion molecules, (b)  $\text{Ca}^{2+}$  requiring,  $\text{Ca}^{2+}$  binding or  $\text{Ca}^{2+}$  regulatory genes, (c) copper-sensitive or copper-transporting proteins and (d) specific regulatory proteins of Rho signaling. Among these functional groups, 23 transcripts involved in cell-cell or cell-matrix interactions are underexpressed in MDA-

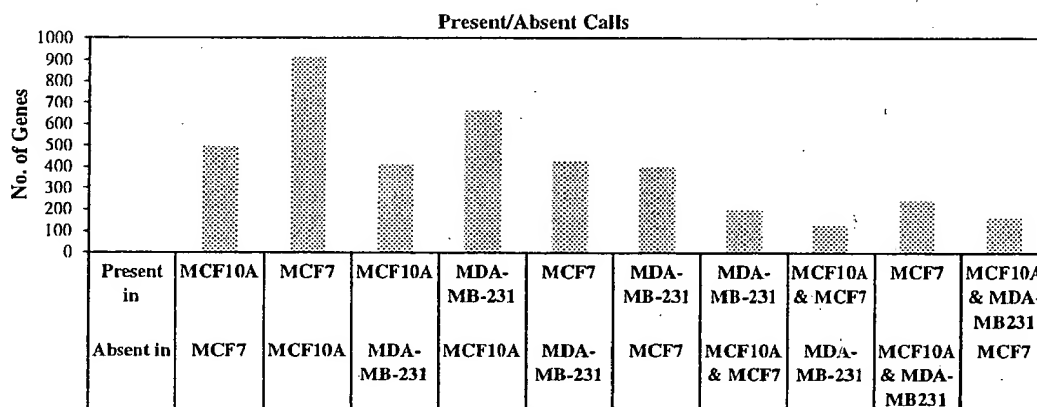
MB-231 cells and 21 transcripts were overexpressed in this cell line (Table 2). The comparison of  $\text{Ca}^{2+}$ -requiring/binding genes indicated downregulation of 26 transcripts and upregulation of 26 transcripts in invasive cells as compared to noninvasive cells (Table 3). While  $\text{Ca}^{2+}$  homeostasis is extensively investigated in human cancers, copper homeostasis is an underexplored area. The alterations in copper homeostasis in breast carcinoma cells were reflected by changes in transcripts corresponding to a variety of copper-binding or copper-sensitive proteins/enzymes (Tables 1 and 4). The deregulation of Rho signaling was evident from changes in various proteins involved in this pathway (Table 4).



**Figure 1** Semiquantitative evaluation of selected transcripts. (a) RNA was isolated from confluent culture dishes containing MCF10A (lanes 1 and 4), MCF7 (lanes 2 and 5) or MDA-MB-231 (lanes 3 and 6) cells. The amount of RNA was first determined spectrophotometrically. Equal amounts of RNA, as determined by absorbance at 260 nm, were amplified with primers specific to actin gene for 18 cycles (lanes 1–3) or 20 cycles (lanes 4–6). The lane containing size markers is labeled as M. (b) A set of primers corresponding to caveolin 2 (lanes 1 and 2), TGF- $\alpha$  (lanes 3 and 4), Moesin (lanes 5 and 6), LOX (lanes 7 and 8), Axl receptor (lanes 9 and 10), RhoD (lanes 11 and 12), S100A13 (lanes 13 and 14), TFF3 (lanes 15 and 16) and Claudin 3 (lanes 17 and 18) were amplified for 32 cycles. Lanes 1, 3, 5, 7, 9, 12, 14, 16 and 18 represent amplified products from MCF7 and lanes 2, 4, 6, 8, 10, 11, 13, 15 and 17 represent MDA-MB-231 cells. The lanes containing PCR products from MDA-MB-231 cells are marked with an asterisk.



**Figure 2** Analysis of selected transcripts by Northern hybridization. The probes specific to CD74 (a), CYR61 (b), SPINT1 (c) and DUSP (d) were labeled with a  $^{32}\text{P}$  nucleotide and hybridized to blots containing RNA from MCF10A (lane 1), MCF7 (lane 2) and MDA-MB-231 (lane 3). The blots were washed stringently and developed as described. The amounts of RNA loaded were normalized as in Figure 1.



**Figure 3** Comparison of cell lines based on the presence or absence of transcripts. The absence or presence of a transcript in the Affymetrix chip was scored by the fluorescence read-out as described in the Materials and methods section.

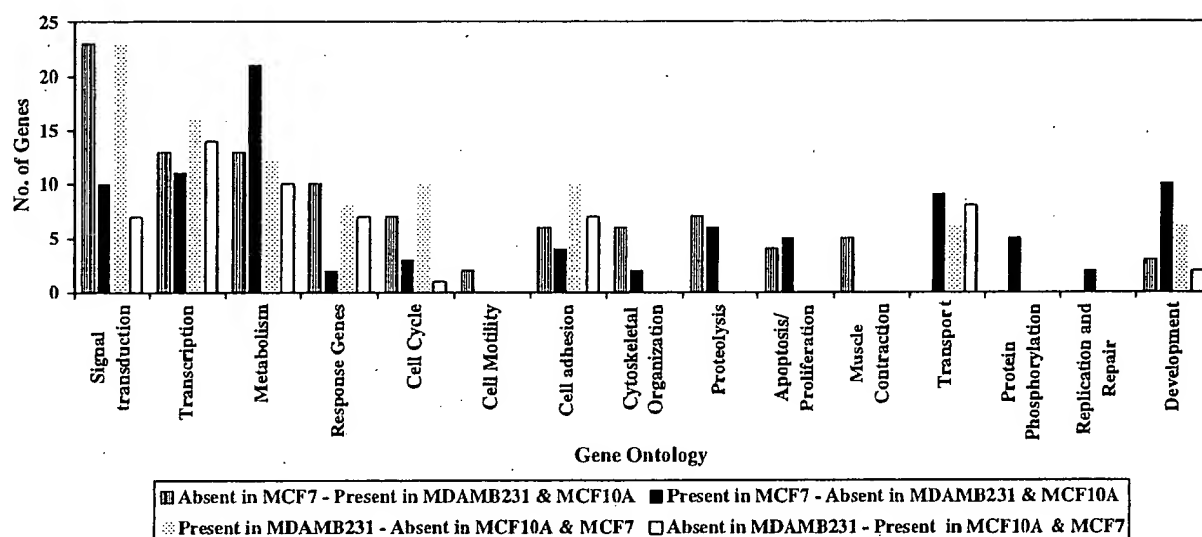


Figure 4 Functional classes of transcripts that differentiate a cell line pair. The transcripts identified as present or absent were classified based on their functional importance.

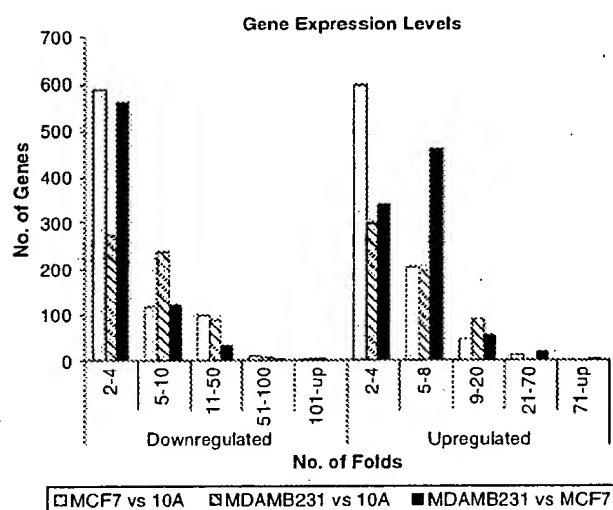


Figure 5 Distribution of altered number of transcripts as a function of fold change. The altered transcripts were categorized in groups based on the magnitude of change in their abundance.

Table 2 Cell adhesion molecules altered in breast carcinoma cells

Upregulated in MDA-MB-231	Downregulated in MDA-MB-231
Cadherin 4	Claudin3
Cadherin 11	Cadherin-1
Catenin	Cadherin-3
Integrin $\alpha$ 6	Cadherin-18
Transmembrane anchor protein	Cadherin, LAG seven pass receptor
Eph B2	Down syndrome cell adhesion
Dystonin	Catenin- $\delta$ 2
Laminin $\beta$ 1	Eph A4
Lamin	Ephrin A4
Filamin B	Annexin A9
Filamin C	Ankyrin 3
Tailin 1	Sarcoglycan
Butyrophilin	Keratin 8
Spectrin- $\alpha$	MAP-7
Spectrin- $\beta$	MAP- $\tau$
Thrombospondin	Plakoglobin
Plastin 3	Plakophilin
Adducin 3 $\gamma$	Discoidin domain receptor
Lamin B receptor	Zona occludens 3
Laminin $\beta$ 2	Periplakin
Lamin A/C	Protocadherin $\alpha$ 9
	Laminin $\gamma$ 2
	Laminin $\alpha$ 3

### Proteome analysis

To identify altered abundance of proteins and relate it to transcript profiles, we characterized the protein profiles of MCF-10A, MCF-7 and MDA-MB-231 cells. Typically, > 300 protein spots could be visualized in silver-stained gels, and there were far fewer protein spots in gels that were stained with Coomassie blue. The comparison of MCF7 or MDA-MB-231 proteins with MCF-10A revealed that MCF-7 had 11 unique protein spots, while MDA-MB-231 had 15 spots that were not seen in MCF-10A. These proteins were either specific to or upregulated in these cell lines. The identity of these proteins is shown in Table 5. Out of these 26 protein spots, only 25 yielded amino-acid sequence. As shown in the table, the list includes proteins involved in stress

response, protein-tagging activities, calcium-binding and calcium homeostasis proteins and some regulatory proteins. Prominent among these changes were proteins involved in calcium homeostasis such as crocalbin, calreticulin, calcyclin and reticulocalbin. The changes in signaling pathways between the two cell lines were indicated by altered levels of Rho GDP dissociation inhibitor 1, an apoptosis/differentiation regulating protein galectin, Myc expression regulator far upstream binding protein-1 and the microtubule regulator protein stathmin. The translation initiation factors 5A and 4H were also selectively upregulated in MDA-MB-231 cells.

Table 3 Calcium binding or sensitive transcripts

Upregulated in MDA-MB-231	Downregulated in MDA-MB-231
Reticulocalbin	S100A8
Dystonin	S100A7
Follistatin like 1	S100A13
Cullin	Tumor associated Ca <sup>2+</sup> signal transducer
Annexin A5	Notch homolog 3
EF hand domain containing-2	PKD2
Hippocalcin like-2	Adenylate cyclase
LDLR	Phospholipase-C
Steroid sulfatase	Chemokine ligand 12
MT-actin crosslinking factor	Ubiquitin-specific protease
Inositol-1,4,5-triphosphate receptor 3	PKC- $\eta$
Sorcin	Ret-protooncogene
Guanine nucleotide-binding protein- $\gamma$	Signal peptide-CUB domain
GAS6	Mannosidase $\alpha$
SWP-70 protein	Bradykinin receptor B2
Jagged 1	Solute carrier family 24
ECGF-containing fibrulin-like ECM	Ca <sup>2+</sup> channel voltage-dependent $\beta$ 3
Transglutaminase 2	Regulator of G protein signaling 17
Thrombin receptor-like 1	Dystrobin- $\alpha$
Plastin 3	Synaptogamin 1
FYN oncogene	Matrix gla protein
PKC- $\alpha$	EF hand domain family member D1
Calmeglin	PK-cyclic AMP dependent
Calpain	ATPase-Ca <sup>2+</sup> transporting
HEG homolog	Calmodulin 1
Cyr61	CaM kinase

Table 4 Altered transcripts involved in copper homeostasis and Rho signaling

Copper-binding proteins	Rho signaling proteins
LOX	Rho3
LOX-1, LOX-2	Rho/Rac GEF 18
SCO cyt oxidase-deficient homolog 2	Rho GEF 12
COX 17 homolog	Ras related C3 botulinum toxin substrate 2
Metallothionein 1E, 1F and 2a	Cdc 42 effector protein 3
Ring finger protein7	Rho GEF 3
Amiloride-binding protein 1	Rho GDP dissociation inhibitor $\beta$
Neurotrypsin/matopsin	RhoD

#### Changes in proliferation characteristics and protein profiles in response to transfection with antisense constructs of selected transcripts

We had observed significant upregulation of transcripts for DNA replication complex protein GINS PSF2, trefoil factor 3, aurora kinase AIK2, protein kinase H11 and secreted protein XAG in MCF7 cells. We reasoned that antisense knockdown of the above genes in MCF7 cells might indicate pathways involved in tumorigenesis and invasiveness.

MCF7 cells were transfected with empty vector pCDNA3.1 or antisense constructs of the above genes. A semiquantitative amplification of pCDNA marker gene by RT-PCR confirmed the presence of the transfected construct in a significant proportion of the cell population. The transfected cells also showed a decrease in the target transcripts as observed by RT-

Table 5 Proteomic profiles of MCF7 and MDA-MB-231 cells as compared to MCF10A cells

MCF7 cells	MDA-MB-231 cells
<i>Cell-cell adhesion/cell-surface receptor proteins</i>	<i>Calcium-binding proteins</i>
Triose-phosphate isomerase	Calcyclin
	Calreticulin
	Crocalbin
	Reticulocalbin
<i>Stress-response proteins</i>	<i>Transcription Factors/Promoter-binding proteins</i>
Hsp27	Far upstream element binding protein-1
Superoxide dismutase	Far upstream element binding protein-2
Peroxioredoxin 2	
<i>Cytoskeletal component/binding proteins</i>	<i>Cell-cell adhesion/cell-surface receptor proteins</i>
Stathmin	Galectin
<i>Cell-cycle regulation and growth/differentiation/apoptosis proteins</i>	<i>ATPase/GTPase/signal transduction/trafficking proteins</i>
Nucleoside diphosphate kinase A	Rho GDP dissociation inhibitor 1
S100C	
<i>Secreted proteins and growth factors</i>	<i>Stress-response proteins</i>
Macrophage migration inhibitory protein	Hsp 70
	Peroxioredoxin 2
<i>Miscellaneous</i>	<i>Cytoskeletal component/binding proteins</i>
Cyt c oxidase VIb	Stathmin
Peptidyl-prolyl <i>cis-trans</i> isomerase	
Ubiquitin	<i>Miscellaneous</i>
	Heterogeneous nuclear ribonucleo protein H
	eIF4H (translation)
	eIF5A (translation)

PCR. The effects of antisense transfections were scored by growth characteristics of the transfectants. The cell proliferation was reduced between 15 and 40% when antisense transfectants were compared to cells transfected with empty vector.

In order to relate decreased proliferation of transfectants to altered proteins, proteomic profiles of transfectants were compared with vector controls. The comparison of protein profiles of cells transfected with empty vector or antisense construct revealed alterations in several proteins for each transfectant (Table 6). These proteins included stress-response proteins, calcium-regulating proteins, translation factors, ubiquitin, proteins of electron transport chain and oxidative phosphorylation, signaling proteins, cytokeratins, actin and actin regulating proteins and general regulatory factors. The number of altered proteins varied between 5 and 15 for various transfections. Peptidyl prolyl *cis-trans* isomerase, calcium-regulating proteins, SOD, galectin, histidine triad protein and PKC substrate were prominent among altered proteins. We performed database searches to identify interactors for all proteins that were altered in transfected cells and identified nearly 350 proteins (data not shown). A significant number of these interacting proteins are involved in transcriptional regulation.

Table 6 Altered proteins in MCF7 cells after antisense knockdown of specific transcripts

After transfection with as-trefoil factor 3 (TFF3)	After transfection with as-protein kinase H11
<b>Calcium-binding proteins</b> Calmodulin <b>Cell-cell adhesion/cell-surface receptor proteins</b> Retinoic acid-binding protein II <b>ATPase-GTPase/signal transduction/trafficking proteins</b> Raf kinase inhibitor <b>Stress-response proteins</b> Hsp27 Peroxiredoxin I <b>Cytoskeletal component/binding proteins</b> Cofilin-nonmuscle isoform <b>Cell-cycle regulation and growth/differentiation/apoptosis proteins</b> Chromobox protein homolog 1 Chromobox protein homolog 3 Translationally controlled tumor protein <b>Miscellaneous</b> 40S ribosomal protein S12 ATP synthase D chain Cancer-associated Sm-like protein Cyt c oxidase polypeptide Va eIF5A (translation) Ferritin heavy chain His triad nucleotide-binding protein Histone H2B.n Peptidyl-prolyl isomerase Proteasome subunit $\beta$ -type 6 RNA-binding protein 8A Thioredoxin Ubiquitin crossreactive protein <b>After transfection with as-aurora kinase AIK2</b> Calcium-binding proteins Calgranulin B Calgranulin A <b>Stress-response proteins</b> Hsp27 Superoxide dismutase <b>Miscellaneous</b> 60S acidic ribosomal protein P2 Peptidyl-prolyl isomerase Ubiquitin-crossreactive protein	<b>Cell-cell adhesion/cell-surface receptor proteins</b> Galectin Retinoic acid-binding protein II <b>ATPase-GTPase/signal transduction/trafficking proteins</b> PKC substrate <b>Stress-response proteins</b> Hsp27 <b>Cytoskeletal component/binding proteins</b> Actin- $\alpha$ , skeletal muscle Cytokeratin 18 <b>Miscellaneous</b> 3OH-acyl CoA dehydrogenase II ATP synthase A chain Cyt c oxidase peptide Va Enhancer of rudimentary homolog Thioredoxin Ubiquinol-cyt C reductase No match <b>After transfection with as-DNA replication complex GINS PSF2</b> <b>Calcium Binding proteins:</b> Calreticulin <b>ATPase-GTPase/signal transduction/trafficking proteins</b> 14-3-3 $\eta$ PKC substrate <b>Cytoskeletal component/binding proteins:</b> Cytokeratin 1 Cytokeratin 18 <b>Cell-cycle regulation and &amp; growth/differentiation/apoptosis proteins</b> Chromobox protein homolog 1 Chromobox protein homolog 3 MAP/MT affinity regulator <b>Miscellaneous</b> ATP synthase D chain Peptidyl-prolyl isomerase Ubiquinol-cyt c reductase <b>After transfection with as-human homolog of XAG</b> <b>Calcium-binding proteins</b> Calreticulin <b>Stress-response proteins</b> Hsp27 <b>Miscellaneous</b> ATP synthase A chain Peptidyl-prolyl isomerase Ubiquitin crossreactive protein No match

## Discussion

The results presented here validate the gene profiles obtained from different expression platforms ranging from subtractive hybridization to gene microarrays and proteomic analysis. The RDA protocol is powerful enough to yield important genes that show significant alterations in their expression between cell lines, and can lead to isolation of full-length cDNAs by using appropriate modifications (Baskaran *et al.*, 1996; Jacob *et al.*, 1997). The detection of RUNX2, variant of RUNX2, EphB6, prion protein, lysyl oxidase and a copper transporter ATP7A transcripts by RDA warrant specific mention. RUNX transcription factors bind specific motifs on target gene promoters and regulate gene expression leading to cell growth, proliferation and

differentiation (Pratap *et al.*, 2003). RUNX2 and its variant have differential repression activity toward the promoter of the cyclin-dependent kinase inhibitor (p21CIP1) (Westendorf *et al.*, 2002). The loss of EphB6 expression due to methylation of its promoter is related to invasiveness of MDA-MB-231 (Fox and Kandpal, 2004; unpublished observations). Lysyl oxidase, a copper-sensitive enzyme, causes oxidative deamination of lysine and hydroxy lysines of collagen to aldehyde forms to stabilize collagen fibrils (Siegel, 1976) that are found in invasive breast carcinoma cells (Akiri *et al.*, 2003). The activation of LOX is dependent on copper that is internalized and then transported to trans golgi network by copper transporter ATP7A (Pase *et al.*, 2004), a protein mutated in Menkes disease (Molier *et al.*, 2005). Prion protein has also been reported as a

chaperone for copper (Jones *et al.*, 2004). Thus, EphB6 can serve as a marker for invasiveness, and LOX and ATP7A may be exploited as relevant targets for therapeutic intervention.

The downregulation of junctional proteins along with inactivation of TIMPs as shown here is in agreement with other reports describing their relationship with invasiveness of carcinoma cells (Johnson, 1991; Kousidou *et al.*, 2004; Shao *et al.*, 2005) and promoter methylation (Costa *et al.*, 2004). As several transcripts coding for junctional proteins are downregulated in invasive cells, we postulate that methylation-specific PCR can be exploited to use these transcripts as biomarkers of tumor cells in general and invasiveness in particular. The changes in cell-cell interaction correlate to cell phenotypes because such interactions influence Rho/Ras signal transduction pathways and *vice versa* (Malliri and Collard, 2003; Nagaraja and Kandpal, 2004; Ridley, 2004), and lend credence to the significance of altered transcripts for Rho and Rho GEFs as presented here.

Early changes in calcium homeostasis as measured by calcium excretion have been reported in breast cancer (Campbell *et al.*, 1983), and altered calcium signaling has been shown in invasive lung carcinoma cells (Amuthan *et al.*, 2002). Prominent among calcium-binding proteins are S-100 protein, a group of intracellular messengers that respond to transient changes in calcium concentration by binding to specific receptors (Marenholz *et al.*, 2004) and regulate cell growth, differentiation and motility, transcription and cell cycle. The S-100 proteins detected in the present study map to chromosome 1q21, a region of genome that is frequently altered in human breast cancer cells (Bieche *et al.*, 1995). Calcium ions act as a second messenger in specific signaling pathways in a variety of cancers (Missiaen *et al.*, 2000) and are known to alter calcineurin to activate transcription factors such as NFATc (Luo *et al.*, 1996).

As dictated by post-transcriptional regulation, protein profiles showed far fewer changes as compared to transcript profiles, and the knockdown of five selected genes in MCF7 cells produced interesting changes in protein profiles. These genes, namely, XAG (secretory *Xenopus laevis* protein), trefoil factors 3, human aurora2 kinase AIK2, protein kinase H11 and DNA replication complex GINS PSF2, have been shown to be estrogen responsive, oncogenic or involved in tumorigenesis (Yu *et al.*, 2001; Fletcher *et al.*, 2003; Katoh, 2003; Warner *et al.*, 2003; Takayama *et al.*, 2003). The antisense constructs of these genes appeared to work as siRNAs as suggested by the reduction in the transcript detected in RT-PCR of RNA isolated from the transfected cells. The involvement of the above transcripts in invasive potential is apparent from the observed upregulation of calcium-binding proteins in transfected MCF7 cells, which is comparable to their levels in MDA-MB-231 cells. The proteins that appear to mediate inhibition of proliferation in antisense-transfected cells include PKC substrate, Raf kinase inhibitor, histidine triad nucleotide-binding protein and peptidyl-prolyl isomerase (Pin1). We believe histidine triad protein effects are

most likely mediated via its interaction with ATM protein. Raf kinase inhibitor (Keller *et al.*, 2004) and ATM (Hall, 2005) have been conclusively linked to transformation of cells, and the activity of Pin1 has been related to p53-mediated signaling pathways (Mantovani *et al.*, 2004; Berger *et al.*, 2005). In this context, p53 activation has also been hypothesized by Cu-SOD prion-like enzyme (Wiseman, 2005). Thus, alterations in copper homeostasis and p53-mediated signaling may be considered as a significant regulatory mechanism in tumorigenesis.

In summary, we have presented here a set of candidate genes that can serve as biomarkers for tumorigenesis and invasiveness, and some of these markers may be used to develop DNA-based diagnostic tests. The alterations in transcripts for copper homeostasis genes suggest copper chelation or inhibition of copper transporter ATP7A as potential targets for therapeutic application. The modulation of RUNX2 splicing variants by chemicals that affect splicing machinery may also be explored as a therapeutic modality. The changes in EphB6 expression, if confirmed in tumor specimens, may have prognostic significance.

## Materials and methods

### Breast cancer cell lines

We used MCF-10A, a cell line established from normal breast, and two breast carcinoma cell lines MCF-7 and MDA-MB-231 that vary in their *in vitro* and *in vivo* invasiveness. All cells were cultured at 37°C/7% CO<sub>2</sub>. MCF-10A cells were grown in 1:1 DMEM:F12 media (Gibco) with 5% horse serum (Gibco), 20 mM HEPES, 10 ng/ml EGF (Invitrogen), 10 ml/l PenStrep-Glutamine (10000 U/ml penicillin, 10000 µg/ml streptomycin and 29.2 mg/ml L-glutamine), 10 µg/ml insulin (Invitrogen), 0.1 µg/ml Cholera Toxin (Sigma) and 500 ng/ml hydrocortisone (Sigma). MCF-7 and MDA-MB-231 cells were grown in DMEM (Gibco) supplemented with 2 mM L-glutamine (Gibco), 1 mM sodium pyruvate (Gibco), 5 ml/l penstrep (5000 U/ml penicillin and 5000 µg/ml streptomycin), and 10% fetal bovine serum (Hyclone).

### Total RNA isolation

RNA was isolated from 85 to 95% confluent 10 cm tissue culture dishes using TRI reagent (Molecular Research Center Inc.) with slight modifications to the recommended protocol. Approximately 10 million cells were mixed with 1.0 ml Tri Reagent, the mixture was extracted with chloroform and the aqueous phase containing RNA was separated. The RNA was precipitated with isopropanol, the pellet washed sequentially with 80 and 100% ethanol, then dried and resuspended in DEPC-treated water. RNA was stored in aliquots at -70°C. The quality of RNA was visualized by running on a formaldehyde gel. The appearance of ribosomal RNA bands indicated that RNA was not degraded during the procedure. The amount of RNA was determined by its absorbance at 260 nm.

### DNAase treatment of total RNA

To remove DNA contamination, 20 µg of RNA (quantified spectrophotometrically) was treated with 500 ng DNAase I;



80 U RNasin (Promega) and 1 mM MgCl<sub>2</sub> in Tris buffer in a total volume of 50  $\mu$ l. The reaction was carried out at 37°C for 1 h and the DNAase inactivated by heating to 65°C for 30 min.

#### Representational difference analysis (RDA)

RDA of cDNAs is a modification of genomic RDA (Lisitsyn *et al.*, 1993). We performed RDA in the following two ways. In one experiment, MCF-7 cDNA was used as a driver and MDA-MB-231 cDNA as a tester. In the second experiment, MCF-7 cDNA was used as a tester and MDA-MB-231 cDNA as a driver. The protocol has been described previously (Jacob *et al.*, 1997). Briefly, first-strand synthesis was carried out using a commercial cDNA synthesis kit as per the manufacturer's protocol. A linker with a *Bgl*II site was ligated to the tester as well as the driver cDNA. A primer specific to one of the linker strands was used to PCR amplify the linker-ligated cDNAs. The linkers were then removed by digesting the cDNA with *Bgl*II and the digested cDNA was gel purified. A second set of unphosphorylated *Bgl*II adaptors was ligated to the tester cDNA only. The tester and driver DNAs were hybridized in a 5  $\mu$ l reaction volume at a ratio of 1:40. After hybridization, the ends of the tester homoduplexes were repaired with Klenow polymerase and 1  $\mu$ l of the reaction mixture was diluted to 100  $\mu$ l. The difference product was obtained by amplifying 1  $\mu$ l of the diluted mixture using the top strand of the ligated adaptor as a primer. The amplified difference product was digested with *Eco*RI and cloned in a pBlueScript vector. Individual clones were picked up and sequenced by Sanger's dideoxy chain termination method. Representative clones were validated by Northern analysis and semiquantitative RT-PCR.

#### Microarray analysis

The GeneChips, Human Genome U133A 2.0, (Affymetrix, Santa Clara, CA, USA) used in this study contained approximately 22 000 probe sets corresponding to 18 400 transcripts and variants, including 14 500 well-characterized human genes.

Total RNA was converted into double-stranded cDNA by using SuperScript II (Invitrogen, Carlsbad, CA, USA) and an oligo-dT primer containing a heel of the T7 RNA polymerase promoter sequence. The reaction mixture containing double-stranded cDNA was extracted with phenol-chloroform, precipitated with ethanol, and dissolved in 12  $\mu$ l RNase-free water. The cDNA was transcribed *in vitro* by using a RNA transcription labeling kit (Enzo Biochem, Farmingdale, NY, USA) with 6  $\mu$ l of double-stranded cDNA in the presence of ATP, CTP, GTP, UTP, bio-11-CTP and bio-16-UTP. The biotinylated RNA was purified by using an affinity column (Qiagen, Valencia, CA, USA) and fragmented randomly, by heating to 95°C in the presence of fragmentation buffer, between sizes of 35 and 200 bases. The GeneChips were hybridized overnight at 45°C in hybridization oven in a solution containing fragmented cRNA, control oligonucleotide B2, 20  $\times$  eucaryotic hybridization controls, herring sperm DNA, acetylated BSA and 2  $\times$  hybridization buffer. The GeneChips were washed and stained with streptavidin-phycoerythrin and the antibody in 2  $\times$  MES stain buffer, acetylated BSA, and optically read at a resolution of 6  $\mu$ m with a Affymetrix GeneChip scanner 3000. Affymetrix MICROARRAY SUITE<sup>3</sup> was used for initial data preparation (generation of .CHP files). Normalization (quantile method) and calculation of signal intensities was performed with the software package RMA from the R project (<http://www.r-project.org/>). For every cell line, three replicates were performed with Affymetrix Gene Chips. The Gene Chip data were used for further calculations after the raw image and

MAS5 analysis revealed a positive quality report. Ratios of average signal intensity (log 2) were calculated for the probe sets between pairs of cell lines and then converted to an average fold change (AFC). Statistical validation was performed on probe sets as described (Yoshida *et al.*, 2004). The statistical method used to assign *P*-values to the fold changes of gene responses is described by Yoshida *et al.* (2004) and is a two-step procedure based on the Benjamini and Yekutieli construction of false discovery rate confidence intervals (FDRCI) (Reiner *et al.*, 2003). Functional annotation of proteins was assigned through Gene Ontology (<http://www.geneontology.org>) or Locuslink (<http://www.ncbi.nlm.nih.gov/LocusLink>) classifications obtained through appropriate public databases.

#### Quantitative RT-PCR

RNA was reverse transcribed with SuperScript II (Invitrogen, Carlsbad, CA, USA) RT by priming with oligodT. The primers specific to validated genes were synthesized from the 3' untranslated region using Primer 3 software. PCR reactions were then performed in triplicates in an I-cycler Thermocycler with optical module (BioRad, Hercules, CA, USA). The amplified products were quantified by reading fluorescence of SybrGreen I (Molecular Probes, Eugene, OR, USA). Average fold changes were calculated by differences in threshold cycles (*C*<sub>t</sub>) between pairs of samples to be compared. *HPRT* gene was used as a control.

#### Semiquantitative RT-PCR

The spectrophotometrically determined concentration of RNA was confirmed by amplifying actin message at different cycles. The cycling conditions that yielded proportional increment of amplified product was used to normalize the RNA concentration. The normalized RNA was used as template to determine relative abundance of transcripts corresponding to clones identified by RDA experiments. The conditions were standardized in the range of cycles that yielded a PCR product for at least one of the pairs of compared RNAs. Such experiments dictated cycles between 30 and 35 to be appropriate to compare abundance of selected transcripts in MCF7 and MDA-MB-231 cells.

#### Northern analysis

The expression pattern of selected transcripts in cell lines was also analysed in Northern blots. RNA (20  $\mu$ m), as determined by spectrophotometer and confirmed by actin amplification, was electrophoresed on a formaldehyde agarose gel. A RNA ladder was used as a size marker. The RNAs were transferred from the gel to a Hybond nylon membrane by capillary transfer. The RNA was fixed onto the membrane by irradiation in a Stratalinker. The blot was hybridized at 65°C for 12–15 h with a radioactive probe and the blot was subsequently washed with 0.1  $\times$  SSC and 1% SDS at 65°C. The hybridized probe was detected by autoradiography.

#### Transfection of MCF-7 cells with antisense constructs

The genes selected on the basis of their upregulation were cloned in antisense orientation in pCDNA3.1 vector (Invitrogen). MCF-7 cells were grown to 70–80% confluence. Approximately 4  $\mu$ g of DNA was transfected into MCF-7 cells by using Lipofectamine 2000. The transfected cells were grown in the presence of G418 (400  $\mu$ g/ml). The transfectants were processed for protein isolation. A control set of cells was transfected with an empty pCDNA3.1. The proteins were analysed by two-dimensional electrophoresis, and altered bands were excised for mass spectrometry.

### Protein isolation

The cultured cells were harvested by trypsinization and centrifuged at 220g for 5 min at 4°C. The cell pellet was washed once with ice-cold 1 × PBS. The proteins were isolated by using a commercial kit (BioRad, Hercules, CA, USA). Briefly, pelleted cells (0.05 ml) were mixed with 0.5 ml ice-cold CPEB solution containing protease inhibitors cocktail (Roche), vortexed and stored on ice for 30 min. The cell suspension was passed through a syringe needle (20 gauge) for 10–20 strokes to ensure complete cell lysis. The cytoplasmic protein fraction was collected by centrifugation at 100g for 10 min at 4°C. The nuclear pellet was washed once again with 0.5 ml CPEB solution. The nuclear pellet was resuspended in 0.75 ml PSB buffer, vortexed briefly and centrifuged at 1000g for 10 min at 4°C, and the supernatant containing nuclear protein was collected into a new tube. The samples were quantified using 2D Quant kit (Amersham Biosciences), aliquoted and stored at –80°C to prevent protein degradation. To reduce streaking, background staining and the other gel artefacts associated with substances contaminating 2D/IEF samples, the samples were cleaned with 2D Clean up kit (Bio Rad, Hercules, CA, USA) before running on the gel.

### Two-dimensional gel electrophoresis

The protein mixtures were separated based on isoelectric points by using commercial pre-cast pH gradient gel strips according to the manufacturer's instructions. The protein sample (175 µg) in 185 µl of sample buffer (8 M urea, 2% CHAPS, 0.2% biolytes, 3/10 ampholytes, 65 mM DTT and 0.002% bromophenol blue) was loaded in the sample loading trays at the end of 11 cm immobilized rehydrated strips (pH 3–10) (Bio Rad, Hercules, CA, USA). Following isoelectric focusing, proteins were reduced and alkylated by successive 15 min treatments with equilibration buffer (6 M urea, 0.375 M Tris-HCl pH 8.8, 2% SDS, 20% glycerol, 2% DTT) and 2.5% (W/V) iodoacetamide, respectively. Proteins were then resolved in the second dimension on 8–16% gradient SDS-PAGE gel (Bio Rad, Hercules, CA, USA). The protein spots were visualized by staining with either silver stain or

Coomassie blue stain. The gel images were compared and bands showing significant (greater than twofold) alterations in intensity were excised and processed for mass spectrometry. Comparisons were made between protein lysates from MCF-10A, MCF-7 and MDA-MB-231 cell lines or between MCF-7 and MCF-7 cells transfected with specific antisense constructs.

### Protein identification by enzymatic digestion followed by mass spectrometry

Prior to performing trypsin digestion, the gel pieces containing protein spots were washed sequentially once with water and twice with acetonitrile. The gel pieces were then allowed to swell in 100 mM ammonium bicarbonate and finally washed with acetonitrile. The washed slices were dried in a Speed Vac concentrator, and subsequently incubated with 20 µl of Promega's autocatalysis-resistant trypsin (12.5 ng/µl in 50 mM ammonium bicarbonate and 5 mM CaCl<sub>2</sub>, pH 8.0) overnight at 37°C. The supernatant (5 µl) from tryptic digest was injected for peptide sequence analysis using on-line capillary liquid chromatography-electrospray ionization-tandem mass spectrometry (LC-MS/MS). The front end HPLC utilized a Dionex (San Francisco, CA, USA) Vydac 300 µm inner diameter × 15 mm C18 column. The linear acetonitrile gradient (3%/min, containing 0.02% TFA) was developed using a Hewlett-Packard 1100 pump operating at 0.1 ml/min, and the flow was split before the injector such that the flow rate through the column was 3 µl/min. Peptides were detected at 215 nm. The in-line mass spectrometer was a ThermoElectron LCQ-DECA instrument operated in data-dependent MS/MS mode, and proteins were identified by searching a nonredundant protein database using the Sequest program.

### Acknowledgements

This research was supported in part by Faculty Research Award, Ames Faculty Award and a grant from the Department of Defense (RPK). We thank Dr M Hamilton for helpful discussions on proteomics.

### References

- Akiri G, Sabo E, Dafni H, Vadasz Z, Kartvelishvily Y, Gan N et al. (2003). *Cancer Res* 63: 1657–1666.
- Alpan RS, Sparvero S, Pardee AB. (1996). *Mol Med* 2: 469–478.
- Amuthan G, Biswas G, Ananadatheerthavarada HK, Vijaya-sarathy C, Shephard HM, Avadhani NG. (2002). *Oncogene* 21: 7839–7849.
- Baskaran N, Kandpal RP, Bhargava AK, Glynn MW, Bale A, Weissman SM. (1996). *Genome Res* 6: 633–638.
- Becker M, Sommer A, Kratzschmar JR, Seidel H, Pohlenz HD, Fichtner I. (2005). *Mol Cancer Ther* 4: 151–168.
- Berger M, Stahl N, Del Sal G, Haupt Y. (2005). *Mol Cell Biol* 25: 5380–5388.
- Bieche I, Champeme MH, Lidereau R. (1995). *Clin Cancer Res* 1: 123–127.
- Campbell FC, Blamey RW, Woolfson AM, Elston CW, Hosking DJ. (1983). *Br J Surg* 70: 202–204.
- Cerosaletti KM, Shapero MH, Fournier RE. (1995). *Genomics* 25: 226–237.
- Costa FF, Verbisck NV, Salim AC, Ierardi DF, Pires LC, Sasahara RM et al. (2004). *Oncogene* 23: 1481–1488.
- Dan S, Tsunoda T, Kitahara O, Yanagawa R, Zembutsu H, Katagiri T et al. (2002). *Cancer Res* 62: 1139–1147.
- Fletcher GC, Patel S, Tyson K, Adam PJ, Schenker M, Loader JA et al. (2003). *Br J Cancer* 88: 579–585.
- Fox BP, Kandpal RP. (2004). *Biochem Biophys Res Commun* 318: 882–892.
- Hall J. (2005). *Cancer Lett* 227: 105–114.
- Jacob AN, Baskaran N, Kandpal G, Narayan D, Bhargava AK, Kandpal RP. (1997). *Somatic Cell Mol Genet* 23: 83–95.
- Johnson JP. (1991). *Metast Rev* 10: 11–22, Review.
- Jones CE, Abdelraheim SR, Brown DR, Viles JH. (2004). *J Biol Chem* 279: 32018–32027.
- Katoh M. (2003). *Int J Mol Med* 12: 3–9.
- Keller ET, Fu Z, Yeung K, Brennan M. (2004). *Cancer Lett* 207: 131–137.
- Khan J, Wei JS, Ringner M, Saal LH, Ladanyi M, Westermann F et al. (2001). *Nat Med* 7: 673–679.
- Kousidou OC, Roussidis AE, Theocharis AD, Karamanos NK. (2004). *Anticancer Res* 24: 4025–4030.
- Lisitsyn N, Lisitsyn N, Wigler M. (1993). *Science* 259: 946–951.
- Luo C, Shaw KT, Raghavan A, Aramburu J, Garcia-Cozar F, Perrino BA et al. (1996). *Proc Natl Acad Sci USA* 93: 8907–8912.

- Malliri A, Collard JG. (2003). *Curr Opin Cell Biol* 15: 583-589.
- Mantovani F, Gostissa M, Collavin L, Del Sal G. (2004). *Cell Cycle* 3: 905-911.
- Marenholz I, Heizmann CW, Fritz G. (2004). *Biochem Biophys Res Commun* 322: 1111-1122.
- Missiaen L, Robberecht W, van den Bosch L, Callewaert G, Parys JB, Wuytack F et al. (2000). *Cell Calcium* 28: 1-21.
- Moller LB, Bukrinsky JT, Molgaard A, Paulsen M, Lund C, Tumer Z et al. (2005). *Hum Mutat* 26: 84-93.
- Nagaraja GM, Kandpal RP. (2004). *Biochem Biophys Res Commun* 313: 654-665.
- Pase L, Voskoboinik I, Greenough M, Camakaris J. (2004). *Biochem J* 378: 1031-1037.
- Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA et al. (2000). *Nature* 406: 747-752.
- Pratap J, Galindo M, Zaidi SK, Vradii D, Bhat BM, Robinson JA et al. (2003). *Cancer Res* 63: 5357-5362.
- Reiner A, Yekutieli D, Benjamini Y. (2003). *Bioinformatics* 19: 368-375.
- Ridley AJ. (2004). *Breast Cancer Res Treat* 84: 13-19.
- Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI et al. (2002). *N Engl J Med* 346: 1937-1947.
- Ross DT, Scherf U, Eisen MB, Perou CM, Rees C, Spellman P et al. (2000). *Nat Genet* 24: 227-235.
- Shao Q, Wang H, McLachlan E, Veitch GI, Laird DW. (2005). *Cancer Res* 65: 2705-2711.
- Siegel RC. (1976). *J Biol Chem* 251: 5786-5792.
- Takayama Y, Kamimura Y, Okawa M, Muramatsu S, Sugino A, Araki H. (2003). *Genes Dev* 17: 1153-1165.
- van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M et al. (2002). *Nature* 415: 530-536.
- Warner SL, Bearss DJ, Han H, Von Hoff DD. (2003). *Mol Cancer Ther* 2: 589-595.
- Westendorf JJ, Zaidi SK, Cascino JE, Kahler R, van Wijnen AJ, Lian JB et al. (2002). *Mol Cell Biol* 22: 7982-7992.
- Wiseman A. (2005). *Med Hypotheses* 65: 32-34.
- Yoshida S, Mears AJ, Friedman JS, Carter T, He S, Oh E et al. (2004). *Hum Mol Genet* 13: 1487-1503.
- Yu YX, Heller A, Liehr T, Smith CC, Aurelian L. (2001). *Int J Oncol* 18: 905-911.

Anuradha Waghay<sup>1</sup>  
Farhana Feroze<sup>1</sup>  
Megan S. Schober<sup>1</sup>  
Fayi Yao<sup>1</sup>  
Chris Wood<sup>2</sup>  
Eric Puravs<sup>2</sup>  
Melissa Krause<sup>2</sup>  
Samir Hanash<sup>2</sup>  
Yong Q. Chen<sup>1,3</sup>

<sup>1</sup>Department of Pathology,  
Wayne State University,  
Detroit, MI, USA

<sup>2</sup>Department of Pediatrics,  
University of Michigan,  
Ann Arbor, MI, USA

<sup>3</sup>Center for Molecular Medicine  
and Genetics, Wayne State  
University, Detroit, MI, USA

## Identification of androgen-regulated genes in the prostate cancer cell line LNCaP by serial analysis of gene expression and proteomic analysis

A common therapy for nonorgan-confined prostate cancer involves androgen deprivation. To develop a better understanding of the effect of androgen on prostatic cells, we have analyzed gene expression changes induced by dihydrotestosterone (DHT) in the androgen responsive prostate cancer line LNCaP, at both RNA and protein levels. Changes at the RNA level induced by DHT were determined by means of serial analysis of gene expression (SAGE), and protein profiling was done by means of quantitative two-dimensional polyacrylamide gel electrophoresis. Among 123 371 transcripts analyzed, a total of 28844 distinct SAGE tags were identified representing 16 570 genes. Some 351 genes were significantly affected by DHT treatment at the RNA level ( $p < 0.05$ ), of which 147 were induced and 204 repressed by androgen. In two independent experiments, the integrated intensity of 32 protein spots increased and 12 decreased at least two-fold in response to androgen, out of a total of 1031 protein spots analyzed. The change in intensity for most of the affected proteins identified could not be predicted based on the level of their corresponding RNA. Our study provides a global assessment of genes regulated by DHT and suggests a need for profiling at both RNA and protein levels for a comprehensive evaluation of patterns of gene expression.

**Keywords:** Androgen / Prostate / LNCaP / Gene expression / Serial analysis of gene expression  
PRO 0137

### 1 Introduction

Androgens affect numerous aspects of prostate biology including development, growth, and maintenance. They also play a critical role in tumorigenesis and progression of prostate cancer. Androgen deprivation is an established treatment modality for prostate cancer. However, the disease eventually progresses into a hormone refractory cancer. Several mechanisms have been identified which may contribute to androgen independence: (1) Mutations in the androgen receptor (AR) lead to ligand-independent activation or promiscuity of the receptor [1, 2]. The ability of the receptor to activate or repress downstream genes can also be affected by mutation. AR mutation is associated with advanced phases of prostate cancer [3, 4]; (2) AR can be activated in a ligand-independent manner by specific growth factors and cytokines [5]; (3) AR gene amplification has been found to occur in

28–30% of tumors that recurred post androgen-ablation therapy [6, 7]. Wallen *et al.* [8] have recently shown that one-third of locally recurrent hormone refractory prostate cancer contain AR gene amplification; (4) Coactivator amplification and corepressor down-regulation have been shown to increase receptor transactivation [9, 10]. Regardless of which pathway(s) is taken by the tumor cells, androgen-regulated genes may ultimately be the key players in the development of hormone refractory cancer. As a step towards a better understanding of the effect of androgen on gene expression, we have undertaken a comprehensive assessment of gene expression changes induced by dihydrotestosterone (DHT) in the androgen responsive prostate cancer cell line LNCaP. An important feature of our study is the parallel assessment of expression changes at both RNA and protein levels.

### 2 Materials and methods

#### 2.1 Prostate cell line and DHT treatment

LNCaP human prostate cancer cell line (American Type Culture Collection, Rockville, MD, USA) was cultured for 3 d in phenol-free RPMI 1640 supplemented with 5% charcoal-stripped fetal bovine serum (FBS) at 37°C. Half of the cultures were then treated with  $10^{-9}$  M DHT for 24 h. Total RNA

**Correspondence:** Dr. Y. Q. Chen, Department of Pathology, Wayne State University, 540 E. Canfield, Detroit, MI 48201, USA  
E-mail: yqchen@med.wayne.edu  
Fax: +1-313-5770057

**Abbreviations:** DHT, dihydrotestosterone; G3PD, glyceraldehyde-3-phosphate dehydrogenase; PSA, prostate specific antigen; RT-PCR, reverse transcription-polymerase chain reaction; SAGE, serial analysis of gene expression

and protein were extracted from untreated and DHT treated cells using TRIzol (Invitrogen, Carlsbad, CA, USA) reagent for RNA and a solubilization cocktail for proteins, consisting of 9 M urea,  $\beta$ -mercaptoethanol and 1% NP-40. For kinetics experiments  $1 \times 10^6$  cells were plated in phenol-free RPMI medium with 5% charcoal-stripped FBS. On day three, cells were treated with 5 mL of media containing  $10^{-9}$  M DHT and total RNA was extracted at different time points from 0–96 h. To determine the direct/indirect effect of androgen on RNA levels, cells were plated as above and treated with 5  $\mu$ g/mL of cyclohexamide in the presence or absence of  $10^{-9}$  M DHT for 24 h and total RNA was extracted.

## 2.2 Serial analysis of gene expression

Serial analysis of gene expression (SAGE) was performed as described previously [11] with the following modifications: ditags were PCR amplified using biotinylated primers and digested with *Nla*III enzyme [12]. Concatemers were heated for 15 min at 65°C and chilled on ice for 10 min before separation on an 8% polyacrylamide gel [13]. The concatemers were then cloned into the *Sph*I site of the pZero vector (Invitrogen). Concatenated tags were screened by PCR using M13 forward and M13 reverse primers. PCR products with inserts greater than 500 bp were isolated and sequenced with M13 forward primer on an automated 3700 DNA sequencer (Perkin-Elmer, Norwalk, CT, USA). For microSAGE, 1  $\mu$ g of total RNA per tube was used for cDNA synthesis in two tubes, with the mRNA Capture Kit (Boehringer Mannheim, Indianapolis, IN, USA). cDNAs were cleaved with *Nla*III, ligated to linkers and then digested with *Bsm*FI enzyme. The released tags were ligated, and processed for the rest of steps as with the standard SAGE protocol.

## 2.3 SAGE data analysis

SAGE tags were extracted using the SAGE software V 4.12 [11]. Tags were matched to the SAGE reliable map (release 10–26–2000) (<http://www.ncbi.nlm.nih.gov/SAGE/>). Due to the fact that some tags map to multiple genes and some genes have multiple tags, SAGE data were analyzed in two different ways: (1) exclusion method: tags that match to multiple genes were discarded. Only tags that match to a single gene were tabulated and composite counts analyzed for their significance; (2) inclusion method: tags that match to multiple genes were counted at 100% toward each gene. All tags were tabulated and composite counts analyzed for their statistical significance. Lists of differentially expressed genes ( $p < 0.05$ ) obtained from the exclusion and inclusion methods were compared, and finally only genes that have a  $p$  value  $< 0.05$  in both lists were considered statis-

tically significant. The total number of genes identified was estimated by  $N_m + (N_{um} - 0.1 N_{um})/3.5$ , where  $N_m$  is the number of genes matched to SAGE tags,  $N_{um}$  is the number of SAGE tags that do not match to known genes or ESTs, with 10% representing the estimated sequencing error per SAGE tag and 3.5, the average number of tags per gene in the SAGE reliable map (release 10–26–2000).

## 2.4 RT-PCR and real-time PCR quantification

For reverse transcription-PCR (RT-PCR), 1  $\mu$ g of total RNA was reverse transcribed into cDNA. One fortieth of cDNA was used for PCR reaction. Samples were collected at different cycles and separated on a 2% agarose gel with ethidium bromide. Image was captured and quantified using NucleoVision 760 Image Workstation (Nucleotech, CA, USA). Amplification curves were generated. Subsequently, RT-PCR was done at cycles within the log phase of amplification.

Real-time quantification was performed in the iCycler (Bio-Rad, Hercules, CA, USA). Briefly, one fortieth of cDNA was used in each reaction. Six reactions were carried out for each gene and three independent experiments were performed. PCR mix comprised of 1X PCR buffer, 1.5 mM  $MgCl_2$ , 0.1 mM dNTP, 200 nM primers (listed below), 0.05 U platinum Taq polymerase (Invitrogen) and 0.1x SYBR green (Molecular Probes, Eugene, OR, USA). PCR was carried out at 94°C for 2 min, and 40 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 30 s. Primers used for real-time PCR quantification: PSA (prostate specific antigen) (Hs.171995, 5'-GGAAATGACCAGGCCAAGAC-3', 5'-CAACCCTGAC C TCACACCTA-3'), SCMH1 (Hs.57475, 5'-GCCTTGACC ACATCACTCCAT-3', 5'-AGGCCTAGGGCTGCAAAAG-3'), and clusterin (Hs.75106, 5'-GCAGGAATACCGCAAAAA GC-3', 5'-GACTCAAGATGCCCCCGTAAG-3'). Standard samples (50, 25, 10, 5, 2.5, 1 and 0.5  $\mu$ L cDNA) and experimental samples were used in real-time quantification PCR. Each sample was run in quadruple reactions. Standard curve ( $C_t = mX + B$ ) was obtained, where  $C_t$  is threshold cycle number,  $X$  is log quantity of target molecules,  $m$  is curve slope and  $B$  is Y-axis intercept value. Number of fold induction or repression for a given target molecule was calculated by  $Q_a/Q_b$ , where  $Q_a$  is the target quantity in experimental sample A and  $Q_b$  is the target quantity in experimental sample B.  $C_{ta} = m \log Q_a + B$  or  $Q_a = 10^{(C_{ta}-B)/m}$ , therefore,  $Q_a/Q_b = 10^{(C_{ta}-C_{tb})/m}$ .

## 2.5 2-D PAGE

The procedure followed was as previously described [14]. Cells were solubilized in 200  $\mu$ L of lysis buffer containing 9.5 M urea (Bio-Rad), 2% NP-40, 2% carrier

ampholytes pH 4–8 (Gallard/Schlessinger, Carle Place, NY, USA), 2%  $\beta$ -mercaptoethanol and 10 mM PMSF. Aliquots containing approximately  $5 \times 10^6$  cells, were applied onto isofocusing gels. IEF was conducted using pH 4 to 8 carrier ampholytes at 700 V for 16 h, followed by 1000 V for an additional 2 h. The 1-D gel was loaded onto the 2-D gel, after equilibration in 125 mM Tris, pH 6.8, 10% glycerol, 2% SDS, 1% DTT and bromophenol blue. For the second dimension separation, a gradient of 11–14% of acrylamide (Serva, Hauppauge, NY, USA) was used. Following 2-D PAGE, proteins were visualized by silver staining of the gels or transferred to an Immobilon-P PVDF membrane (Millipore, Bedford, MA).

## 2.6 Protein identification by mass spectrometry

The 2-D gels were silver stained by successive incubations in 0.02% sodium thiosulfate for 2 min, 0.1% silver nitrate for 40 min and 0.014% formaldehyde plus 2% sodium carbonate. The proteins of interest were excised from the 2-D gels and destained for 5 min in 15 mM potassium ferricyanide and 50 mM sodium thiosulfate as described [15]. Following three washes with water, the gel pieces were dehydrated in 100% acetonitrile for 5 min and dried for 30 min in a vacuum centrifuge. Digestion was performed by addition of 100 ng of trypsin (Promega, Madison, WI, USA) in 200 mM ammonium bicarbonate or by addition of 100 ng of the endoprotease Glu-C (Promega) in 100 mM ammonium bicarbonate. The Lys-C digestion was performed with 500 ng of the endoprotease Lys-C (Roche, Mannheim, Germany) in 100 mM Tris-HCl, pH 9. Following enzymatic digestion overnight at 37°C, the peptides were extracted twice with 50  $\mu$ L of 60% acetonitrile/1% TFA. After removal of acetonitrile by centrifugation in a vacuum centrifuge, the peptides were concentrated by using pipette tips C18 (Millipore).

Analyses were performed primarily using a PerSeptive Biosystem MALDI-TOF Voyager-DE mass spectrometer (Framingham, MA, USA), operated in delayed extraction mode. Peptide mixtures were analyzed using a saturated solution of cyano-4-hydroxycinnamic acid (CHA) (Sigma, St. Louis, MO, USA) in acetone containing 1% TPA (Sigma). Peptides were selected in the mass range of 800–4000 Da. Spectra were calibrated using calibration mixture 2 of the Sequazyme peptide mass standards kit (PerSeptive Biosystems). The search program MS-Fit, developed by the University of California at San Francisco (<http://prospector.ucsf.edu>), was used for searches in the NCBI database. Search parameters were as follows: maximum allowed peptide mass error of 400 ppm, consideration of one incomplete cleavage per peptide and

pH range between 4 and 8. MALDI-TOF mass spectrometry was also used for molecular weight determination as described [16]. In some cases, the amino acid sequence of some peptides of interest was determined by ESI MS analysis.

## 3 Results

### 3.1 SAGE analysis of the effect of androgen on gene expression

SAGE libraries were generated from LNCaP cells cultured in the presence or absence of DHT. A total of 123 371 tags were generated of which 62 878 were from the LNCaP cell line and 60 493 from the LNCaP cells treated with DHT for 24 h (Table 1). Sequence analysis identified a total of 28 844 distinct tags representing 16 570 genes, 11 243 from LNCaP and 12 203 from DHT treated cells. A total of 351 transcripts were differentially expressed at a significant level ( $p < 0.05$ ). Eighty-seven percent of transcripts matched GenBank entries; 79% corresponded to known sequences and 8% to ESTs. RNA levels for 147 genes were increased and 204 genes were decreased after DHT treatment (Table 2). Therefore, at the RNA level more genes were repressed than activated by androgens. Of these androgen-regulated genes, 149 were changed  $\geq$  five-fold by DHT treatment (Table 3).

Table 1. Summary of SAGE analysis in LNCaP cells

Sample	Total transcripts	Distinct tags	Number of genes
– DHT	62 878	17 050	11 243
+ DHT	60 493	18 510	12 203
Total	123 371	28 844	16 570

– DHT: LNCaP cells without dihydrotestosterone; + DHT: LNCaP cells treated with  $10^{-9}$  M dihydrotestosterone for 24 h

Table 2. DHT regulated genes in LNCaP cells

Genes	LNCaP vs LNCaP + DHT (gene #)	
	$p < 0.05$	$p < 0.01$
Known genes	277	147
ESTs	29	8
No match	45	13
Total	351	168
Up-regulated by DHT (24 h)	147	65
Down-regulated by DHT (24 h)	204	103

Table 3. List of candidate genes that are regulated by DHT ( $\geq 5$  fold)

UGD or Tag	CAP	DHT	Change	P value	Description
283305	1	82	Induced	0	immunoglobulin heavy constant alpha 1
183752	0	42	Induced	0	microseminoprotein, beta-
140	1	44	Induced	0	immunoglobulin heavy constant gamma 3 (Gm marker)
75415	18	95	Induced	0	beta-2-microglobulin
84298	0	17	Induced	1.00E-05	CD74 antigen (invariant polypeptide of major histocompatibility complex, class II antigen-associated)
9615	0	16	Induced	1.00E-05	myosin regulatory light chain 2, smooth muscle isoform
77443	0	16	Induced	1.00E-05	actin, gamma 2, smooth muscle, enteric
1119	0	15	Induced	2.33E-05	nuclear receptor subfamily 4, group A, member 1
84753	23	2	Repressed	3.00E-05	KIAA0246 protein
75777	0	14	Induced	3.66E-05	transgelin
	0	14	Induced	3.66E-05	immunoglobulin J polypeptide, linker protein for immunoglobulin alpha and mu polypeptides
76325					
1852	1	17	Induced	4.33E-05	acid phosphatase, prostate
78465	0	13	Induced	9.00E-05	v-jun avian sarcoma virus 17 oncogene homolog
78344	1	16	Induced	1.20E-04	myosin, heavy polypeptide 11, smooth muscle
83006	19	2	Repressed	1.50E-04	CGI-139 protein
263812	15	1	Repressed	3.19E-04	nuclear distribution gene C ( <i>A. nidulans</i> ) homolog
TACGGGGATA	0	11	Induced	4.20E-04	Novel
GCCTGGGTGG	11	0	Repressed	6.63E-04	Novel
GACTGACACT	16	2	Repressed	8.40E-04	Novel
284296	0	10	Induced	8.60E-04	Homo sapiens SURF-4 mRNA, complete cds
75105	13	1	Repressed	1.19E-03	emopamil-binding protein (sterol isomerase)
128075	10	0	Repressed	1.30E-03	ESTs
154162	15	2	Repressed	1.43E-03	ADP-ribosylation factor-like 2
6895	1	12	Induced	1.49E-03	actin related protein 2/3 complex, subunit 3 (21 kD)
143240	17	3	Repressed	1.63E-03	ESTs
211582	0	9	Induced	1.66E-03	myosin, light polypeptide kinase
180266	0	9	Induced	1.66E-03	tropomyosin 2 (beta)
93002	14	2	Repressed	2.53E-03	ubiquitin carrier protein E2-C
17883	14	2	Repressed	2.53E-03	protein phosphatase 1G (formerly 2C), magnesium-dependent, gamma isoform
126023	14	2	Repressed	2.53E-03	ESTs, Highly similar to NTC1_HUMAN NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1 PRECURSOR
69469	2	13	Induced	2.92E-03	dendritic cell protein
77899	2	13	Induced	2.92E-03	tropomyosin 1 (alpha)
285501	2	13	Induced	3.07E-03	Human rearranged immunoglobulin lambda light chain mRNA
119209	0	8	Induced	3.07E-03	insulin-like growth factor binding protein 7
173043	0	8	Induced	3.07E-03	metastasis-associated 1-like 1
75866	0	8	Induced	3.07E-03	dimethylarginine dimethylaminohydrolase 1
171695	0	8	Induced	3.07E-03	dual specificity phosphatase 1
TACGGGGATT	0	8	Induced	3.34E-03	Novel
172791	13	2	Repressed	4.19E-03	ubiquitously-expressed transcript
13561	13	2	Repressed	4.19E-03	ESTs, Weakly similar to dJ37E16.5 [ <i>H. sapiens</i> ]
98260	8	0	Repressed	4.34E-03	ESTs
CATAAGACTT	8	0	Repressed	4.64E-03	Novel
TACGGGGACA	2	12	Induced	5.76E-03	Novel
180034	12	2	Repressed	6.91E-03	cleavage stimulation factor, 3' pre-RNA, subunit 3, 77kD
5753	12	2	Repressed	6.91E-03	inosito(myo)-1(4)-monophosphatase 2
278941	10	1	Repressed	6.94E-03	PRO0628 protein
19500	0	7	Induced	7.25E-03	nuclear localization signal deleted in velocardiofacial syndrome

Table 3. Continued

UGD or Tag	CAP	DHT	Change	P value	Description
9006	0	7	Induced	7.25E-03	VAMP (vesicle-associated membrane protein)-associated protein A (33kD)
26471	0	7	Induced	7.25E-03	<i>Homo sapiens</i> clone HQ0692
AACTGCTGGC	12	2	Repressed	7.63E-03	Novel
GCTGACCGTC	7	0	Repressed	8.59E-03	Novel
CCCCCCTGTC	7	0	Repressed	8.59E-03	Novel
243901	2	11	Induced	8.82E-03	<i>Homo sapiens</i> cDNA FLJ20738 fis, clone HEP08257
119000	1	9	Induced	9.12E-03	actinin, alpha 1
78596	1	9	Induced	9.12E-03	proteasome (prosome, macropain) subunit, beta type, 5
171955	7	0	Repressed	9.61E-03	trophinin associated protein (tastin)
180545	7	0	Repressed	9.61E-03	<i>Homo sapiens</i> mRNA for hypothetical protein (TR2/D15 gene)
77719	7	0	Repressed	9.61E-03	gamma-glutamyl carboxylase
4877	7	0	Repressed	9.61E-03	CGI-51 protein
22795	7	0	Repressed	9.61E-03	ESTs
79335	7	0	Repressed	9.61E-03	<i>Homo sapiens</i> SWI/SNF-related, matrix-associated, actin-dependent regulator of chromatin D1 (SMARCD1) mRNA
119177	9	1	Repressed	1.26E-02	ADP-ribosylating factor 3
CGGGAGCACC	9	1	Repressed	1.29E-02	Novel
182217	0	6	Induced	1.32E-02	succinate-CoA ligase, ADP-forming, beta subunit
75106	0	6	Induced	1.32E-02	clusterin (complement lysis inhibitor, SP-40, 40, sulfated glycoprotein 2, testosterone-repressed prostate message 2)
256311	0	6	Induced	1.32E-02	granin-like neuroendocrine peptide precursor
103180	0	6	Induced	1.32E-02	DC2 protein
TACGGGGATG	1	8	Induced	1.70E-02	Novel
AAACAAATCA	2	10	Induced	1.70E-02	Novel
136644	2	10	Induced	1.71E-02	CS box-containing WD protein
8036	2	10	Induced	1.71E-02	glioblastoma overexpressed
74284	2	10	Induced	1.71E-02	ESTs, Weakly similar to Similar to <i>S. cerevisiae</i> hypothetical protein L3111
ATGGCTGATC	6	0	Repressed	1.75E-02	Novel
ATCACTGCCC	6	0	Repressed	1.75E-02	Novel
ACATCATCAG	6	0	Repressed	1.75E-02	Novel
CCAGTCCAAG	6	0	Repressed	1.75E-02	Novel
54842	1	8	Induced	1.78E-02	ESTs
1526	1	8	Induced	1.78E-02	ATPase, Ca++ transporting, cardiac muscle, slow twitch 2
7911	1	8	Induced	1.78E-02	KIAA0323 protein
227400	1	8	Induced	1.78E-02	mitogen-activated protein kinase kinase kinase 3
77269	1	8	Induced	1.78E-02	guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 2
7943	6	0	Repressed	1.81E-02	RPB5-mediating protein
278544	6	0	Repressed	1.81E-02	acetyl-Coenzyme A acetyltransferase 2 (acetoacetyl Coenzyme A thiolase)
23111	6	0	Repressed	1.81E-02	phenylalanine-tRNA synthetase-like
31442	6	0	Repressed	1.81E-02	RecQ protein-like 4
38628	6	0	Repressed	1.81E-02	hypothetical protein
77422	6	0	Repressed	1.81E-02	proteolipid protein 2 (colonic epithelium-enriched)
211973	6	0	Repressed	1.81E-02	homolog of Yeast RRP4 (ribosomal RNA processing 4), 3'-5'-exoribonuclease
171075	6	0	Repressed	1.81E-02	replication factor C (activator 1) 5 (36.5kD)
89781	6	0	Repressed	1.81E-02	upstream binding transcription factor, RNA polymerase I
154149	6	0	Repressed	1.81E-02	apurinic/apyrimidinic endonuclease(APEX nuclease)-like 2 protein



Table 3. Continued

UGD or Tag	CAP	DHT	Change	P value	Description
279 772	8	1	Repressed	2.28E-02	brain specific protein
205 091	8	1	Repressed	2.28E-02	ESTs, Weakly similar to WW domain binding protein 11 ( <i>M. musculus</i> )
75 658	8	1	Repressed	2.28E-02	phosphorylase, glycogen; brain
GGGCAGCTGT	8	1	Repressed	2.37E-02	Novel
105440	0	5	Induced	2.43E-02	hepatocyte nuclear factor 3, alpha
118244	0	5	Induced	2.43E-02	protein phosphatase 2, regulatory subunit B (B56), delta isoform
82389	0	5	Induced	2.43E-02	CGI-118 protein
31638	0	5	Induced	2.43E-02	restin (Reed-Steinberg cell-expressed intermediate filament-associated protein)
12013	0	5	Induced	2.43E-02	ATP-binding cassette, sub-family E (OABP), member 1
12797	0	5	Induced	2.43E-02	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 16
7736	0	5	Induced	2.43E-02	hypothetical protein
274479	0	5	Induced	2.43E-02	NME7
153138	0	5	Induced	2.43E-02	origin recognition complex, subunit 5 (yeast homolog)-like
16034	0	5	Induced	2.43E-02	ESTs
270072	0	5	Induced	2.43E-02	ESTs
187035	0	5	Induced	2.43E-02	ESTs
140452	0	5	Induced	2.43E-02	cargo selection protein (mannose 6 phosphate receptor binding protein)
182265	0	5	Induced	2.43E-02	keratin 19
19762	0	5	Induced	2.43E-02	ESTs, Weakly similar to unknown ( <i>D. melanogaster</i> )
317	0	5	Induced	2.43E-02	topoisomerase (DNA)
6236	0	5	Induced	2.43E-02	ESTs
GCTGGAGCCT	5	0	Repressed	3.03E-02	Novel
CCAGTGCTCA	5	0	Repressed	3.03E-02	Novel
ACCCTACATA	5	0	Repressed	3.03E-02	Novel
GGGGAAATCT	5	0	Repressed	3.03E-02	Novel
ACTGGTACTG	5	0	Repressed	3.03E-02	Novel
GCTCCGGTGT	5	0	Repressed	3.03E-02	Novel
ACAGTGGTGA	5	0	Repressed	3.03E-02	Novel
7869	1	7	Induced	3.04E-02	lysophosphatidic acid acyltransferase-delta
12101	1	7	Induced	3.04E-02	hypothetical protein
266940	1	7	Induced	3.04E-02	t-complex-associated-testis-expressed 1-like 1
6196	1	7	Induced	3.04E-02	integrin-linked kinase
366	1	7	Induced	3.04E-02	6-pyruvoyltetrahydropterin synthase
173611	1	7	Induced	3.04E-02	NADH dehydrogenase (ubiquinone) Fe-S protein 2 (49kD) (NADH-coenzyme Q reductase)
102469	1	7	Induced	3.04E-02	putative nuclear protein
78825	1	7	Induced	3.04E-02	matrin 3
284465	1	7	Induced	3.04E-02	ESTs
30738	1	7	Induced	3.04E-02	hypothetical protein FLJ10407
78687	1	7	Induced	3.04E-02	neutral sphingomyelinase (N-SMase) activation associated factor
92381	1	7	Induced	3.04E-02	nudix (nucleoside diphosphate linked moiety X)-type motif 4
242039	5	0	Repressed	3.26E-02	EST
4766	5	0	Repressed	3.26E-02	DKFZP586O0120 protein
283109	5	0	Repressed	3.26E-02	hypothetical protein DKFZp762L1710
192853	5	0	Repressed	3.26E-02	ubiquitin-conjugating enzyme E2G 2 (homologous to yeast UBC7)
153678	5	0	Repressed	3.26E-02	reproduction 8

Table 3. Continued

UGD or Tag	CAP	DHT	Change	P value	Description
251317	5	0	Repressed	3.26E-02	EST
279623	5	0	Repressed	3.26E-02	selenoprotein X
150319	5	0	Repressed	3.26E-02	ESTs
102456	5	0	Repressed	3.26E-02	survival of motor protein interacting protein 1
284250	5	0	Repressed	3.26E-02	AD-003 protein
251871	5	0	Repressed	3.26E-02	CTP synthase
270480	7	1	Repressed	3.92E-02	ESTs, Weakly similar to ALU5_HUMAN ALU SUBFAMILY SC SEQUENCE
26655	7	1	Repressed	3.92E-02	glucose-6-phosphatase, transport (glucose-6-phosphate) protein 1
8118	7	1	Repressed	3.92E-02	leukotriene A4 hydrolase
CTCCGCCGGC	7	1	Repressed	4.47E-02	Novel
AGGAAATGCT	7	1	Repressed	4.47E-02	Novel
GCTGACCGAGG	7	1	Repressed	4.47E-02	Novel
CCCATAGTCC	7	1	Repressed	4.47E-02	Novel

CAP: LNCaP cells without dihydrotestosterone

DHT: LNCaP cells treated with  $10^{-9}$  dihydrotestosterone for 24 h

UGD: unigene ID

Numbers in CAP and DHT columns are the number of tags observed in respective samples

### 3.2 Confirmation of SAGE data

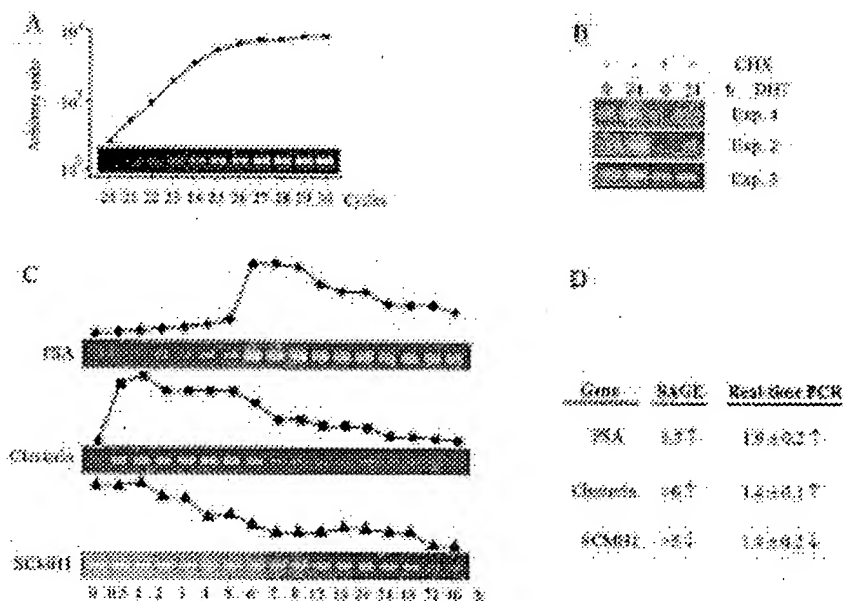
To confirm the differential expression pattern at the RNA level, semiquantitative RT-PCR assay was performed on a group of selected genes. One  $\mu$ g of total RNA was reverse transcribed into cDNA with oligo dT. cDNA was amplified by PCR for various cycles. PCR products were separated on agarose gels and quantified by densitometry. PCR amplification curves were plotted for each gene, and data within the logarithmic phase of amplification were used for quantification (Fig. 1A, B and C). The reproducibility of gene regulation by androgen was confirmed in three independent experiments by monitoring the induction of PSA in the presence or absence of cyclohexamide (Fig. 1B). The kinetics of gene induction or repression was determined for PSA (Hs.171995), clusterin (Hs.75106) and SCMH1 (Hs.57475) genes. Typical examples of kinetics are shown in Figure 1C. PSA was induced at 4–6 h, peaked between 6–20 h, and gradually declined after 20 h post-treatment of DHT. Clusterin was induced within 0.5–1 h and gradually declined after 6–12 h. SCMH1 was repressed 2–4 h post-treatment. Expression of PSA, clusterin, and SCMH1 was quantified by real-time PCR (Fig. 1D).

### 3.3 Changes in protein expression induced by DHT

Protein lysates of LNCaP cells cultured in the presence of DHT for 72 h and cells cultured in parallel in the absence of DHT were subjected to 2-D PAGE (Fig. 2). Following sil-

ver staining, gels were digitized prior to computer-based matching and quantitative analysis, as previously described [17]. Of a total of 1031 protein spots matched between 2-D patterns of DHT treated and untreated cells, a set of 32 protein spots increased in intensity by at least two-fold in two independent experiments. Likewise, a set of 12 protein spots decreased in intensity by at least two-fold in these two independent experiments. A two-fold change in integrated intensity by silver staining represents on average a three-fold change in amount of protein, based on prior quantitative studies [17].

The 44 protein spots that changed in intensity as a result of DHT treatment were excised from the gels, digested with trypsin, and subsequently analyzed by MALDI-TOF MS. The resulting spectra were used to identify the proteins using the MS-Fit search program. Of the 44 spots excised from the gels, 29 were identified without ambiguity and consisted of 21 up-regulated and eight down-regulated proteins (Table 4). The identified proteins represented a heterogeneous group that included cytoskeletal proteins (e.g. tropomyosin,  $\alpha$  tubulin), metabolic enzymes (e.g. adenine phosphoribosyl transferase,  $\beta$  1,4 galactosyl transferase, galactocerebrosidase), and the products of previously described androgen responsive genes (e.g. SRY, selenium binding protein) [18, 19]. Specific antibodies confirmed the identification based on MS for all proteins analyzed by Western blotting, which included  $\alpha$  tubulin, myosin light chain isoforms, nucleoside diphosphate kinase A, glyceraldehyde 3-phosphate dehydrogenase (G3PD), and tropomyosin (data not shown).



**Figure 1.** Kinetics and quantification of androgen-regulated genes. Total RNA from untreated and DHT treated LNCaP cells was reverse transcribed. cDNA was PCR amplified for various cycles. PCR amplification curves were plotted for each gene. A cycle number within the logarithmic phase of amplification was selected for semi-quantitative analysis. cDNA was also used for real-time quantitative PCR. (A) An example of RT-PCR amplification of PSA at cycles between 20–30. (B) RT-PCR of PSA in three independently treated LNCaP samples. h: hours; 0: no DHT; 24: treated with  $10^{-6}$  M DHT for 24 h; –: without cycloheximide; +: with 5  $\mu$ g/mL of cycloheximide for 24 h. (C) An example of three independent kinetics experiments for PSA, clusterin and SCMH1. (D) Comparison of SAGE results with that of real-time quantitative PCR for PSA, clusterin and SCMH1. ↑: induction; ↓: repression.

without cycloheximide; +: with 5  $\mu$ g/mL of cycloheximide for 24 h. (C) An example of three independent kinetics experiments for PSA, clusterin and SCMH1. (D) Comparison of SAGE results with that of real-time quantitative PCR for PSA, clusterin and SCMH1. ↑: induction; ↓: repression.

Corresponding SAGE data were available for most of the proteins affected by DHT treatment that were identified. It was therefore of interest to determine if the changes observed at the protein levels were matched by concordant changes at the mRNA level. Remarkably, for most of the proteins identified, there was no appreciable concordant change at the RNA level (Table 4).

#### 4 Discussion

We have compared the gene expression profile of LNCaP, an androgen responsive prostate cancer cell line, with or without androgen (DHT) treatment. Approximately 350 of 16570 expressed genes detected at the RNA level were affected by dihydrotestosterone at the  $p < 0.05$  level. The DHT responsive genes included known genes, ESTs, and novel genes. As expected, we saw an induction of genes that are well known to be regulated by androgens. For instance, we found a 1.7-fold induction in kallikrein 3/PSA, a 7.6-fold induction in prostatic kallikrein 2, and a 15.7-fold induction in prostatic acid phosphatase by DHT in LNCaP cells. We also saw a five-fold induction in NKX3.1/NKX3A and three-fold in fatty acid synthase; two previously identified androgen-regulated genes [20, 21]. Interestingly  $\beta$ -microseminoprotein, reported to be reduced or lost in prostate tumor [22], was up-regulated more than 40-fold in

LNCaP cells by DHT. More significantly, our data indicate that genes involved in a variety of tumor cell functions such as growth, apoptosis, and metastasis, are directly or indirectly regulated by androgens. In addition, it is noteworthy that both  $\beta$ -actin and G3PD, the two most frequently used loading controls, were up-regulated approximately two-fold by DHT ( $p < 0.01$ ). Thus, these two genes may not be appropriate controls in some experiments.

Serial analysis of androgen-regulated gene expression provides us with a list of candidate genes. However, many factors such as the dose of DHT and the time of cell exposure to DHT and other unknown experimental variations will affect the level of gene expression. Therefore, it is important to confirm the SAGE results with alternative methods. We performed semiquantitative RT-PCR (Fig. 1B) on approximately 20 genes and real-time quantitative PCR on 10 genes in at least three independently DHT treated LNCaP samples. We noticed experimental variations in every gene determined. Some genes such as PSA (Fig. 1B) have smaller variations and some have larger variations. In addition, the fold of alteration in expression identified by SAGE is different for some genes compared to that identified by real-time quantitative PCR. This is probably due to the technical limitation of SAGE. Serial analysis of gene expression is highly quantitative for genes whose tags are detected at large numbers.

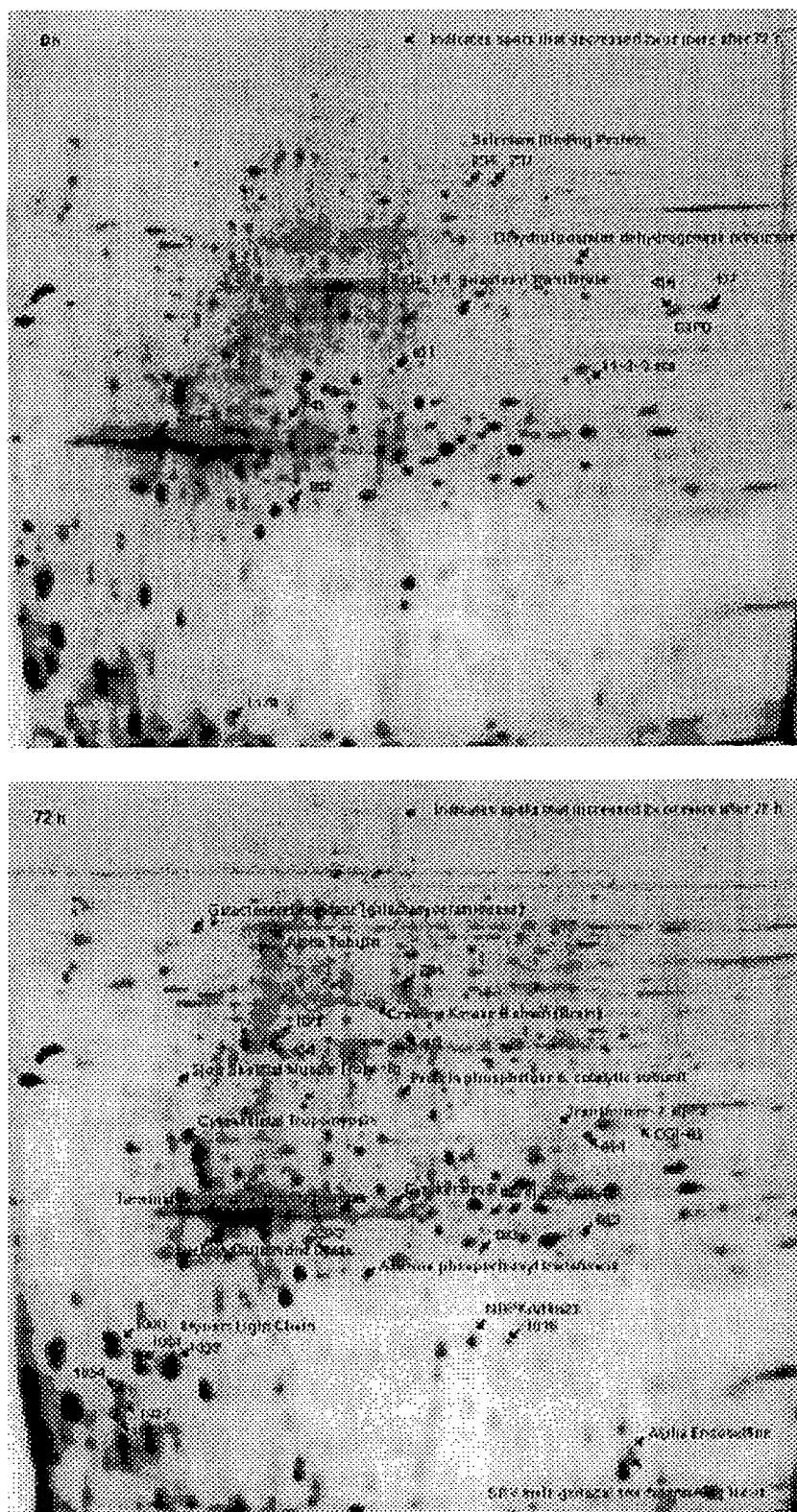


Figure 2. 2-D PAGE patterns of untreated (A) and LNCaP cells treated with DHT for 72 h (B). Identified proteins that were reduced in levels following DHT treatments are shown in (A) and induced proteins are shown in (B). Numbered spots represent either proteins that were changed in level that have not been identified or isoforms of identified proteins.

Table 4. List of proteins that are altered by DHT ( $\geq 2$ -fold)

Protein name or spot number	Unigene ID	Protein		Change in protein	mRNA		Change in mRNA
		0 h	72 h		0 h	24 h	
Adenine phosphoribosyl transferase	28 914	0.448	1.084	↑	36	18	↓
Alpha Endosulfine	111 680	0.066	0.679	↑	2	5	↑
Alpha Tubulin	278 242	0.924	1.932	↑	48	38	↓
Nucleophosmin, B23	173 205	0.256	0.522	↑	16	5	↓
CGI-83	118 554	0.063	0.291	↑	2	3	↑
Creatine Kinase B chain (brain)	173 724	0.508	1.142	↑	59	32	↓
Cytokeratin 8	242 463	0.134	0.393	↑	35	40	↑
Cytoskeletal tropomyosin, NTRK1	85 844	0.414	1.174	↑	14	18	↑
Galactocerebrosidase	273	0.095	0.405	↑	3	3	=
Lactoyl Glutathione Lyase	75 207	0	1.256	↑	27	22	↓
Myosin Light Chain (1000)	77 385	0.724	6.102	↑	118	88	↓
Myosin Light Chain (1001)	77 385	0.6	3.212	↑	118	88	↓
Myosin Light Chain (1032)	77 385	2.573	5.694	↑	118	88	↓
Nm23	118 638	0.583	1.978	↑	35	25	↓
Protein phosphatase 6, catalytic	279 563	0.287	0.737	↑	2	2	=
Protein	7 016	0	0.448	↑	2	5	↑
Slow Skeletal Muscle Troponin	84 673	0.067	0.195	↑	1	0	↓
SRY male gonadal sex determining	1 992	0	0.735	↑	ND	ND	
Synovial sarcoma, X breakpoint 2	289 105	0.225	0.593	↑	ND	ND	
Terminal deoxynucleotidyl	272 537	0.392	0.995	↑	ND	ND	
Transformer-2 alpha	119 523	0.074	0.201	↑	7	13	↑
284		0.173	0.461	↑			
402		0	0.33	↑			
456		0.154	0.314	↑			
664		0.297	0.629	↑			
807		0	0.377	↑			
863		0.246	0.665	↑			
903		0.271	0.632	↑			
1018		0	0.725	↑			
1054		0.47	1.394	↑			
1057		0.451	0.959	↑			
1082		0.609	1.738	↑			
14-3-3 eta	75 544	0.232	0.114	↓	1	5	↑
Beta-1,4-galactosyl transferase	158 540	1.306	0.707	↓	ND	ND	
Cox11	241 515	1.643	1.453	↓	3	0	↓
Dihydrolipoamide dehydrogenase	74 635	0.844	0.132	↓	9	8	↓
G3PD (456)	169 476	1.191	0.279	↓	75	137	↑
G3PD (457)	169 479	2.051	0.336	↓	75	137	↑
Selenium Binding Protein (234)	7 833	0.411	0.201	↓	26	31	↑
Selenium Binding Protein (237)	7 833	0.574	0.22	↓	26	31	↑
651		0.31	0.115	↓			
745		0.21	0.067	↓			
892		0.674	0.454	↓			
1178		0.933	0.305	↓			

Numbers in protein are the intensity of protein spots

Numbers in mRNA columns are the number of SAGE tags detected

ND: not detected

↑: increase; ↓: decrease; =: equal

When the tag number detected for a given gene is close to zero, the quantitative nature of SAGE is compromised. For instance, in approximately 60 000 transcripts of each sample, the PSA tag was detected 42 times in LNCaP and 70 times in LNCaP treated with DHT. However, the tag for clusterin was detected 0 times and 6 times, and the tag for SCMH1 5 times and 0 times in the corresponding samples. Therefore, the SAGE data of PSA is more reliable than that of clusterin and SCMH1. Indeed, SAGE data of PSA is virtually identical to that of real-time quantitative PCR, whereas the SAGE data of clusterin and SCMH1 differ from that of quantitative PCR (Fig. 1D).

PSA is probably the best-known androgen-regulated gene. We were surprised to see only 1.7-fold induction by DHT. However, kinetics experiments indicate that PSA is induced at 4–6 h, peaked between 6–20 h, and gradually declined 20 h post-treatment of DHT. Since the SAGE experiment was done in samples treated for 24 h with DHT, PSA mRNA level was likely to have declined from its peak level. Indeed, we could detect 5–10 fold induction of PSA at 6–8 h post-treatment. Clusterin was reported as an androgen-repressed gene in the rat prostate [23]. Recent evidence indicates that clusterin may not be directly androgen-repressed, but regulated by apoptotic stimuli [24]. Our results suggest that clusterin was induced within 0.5–1 h, gradually declined after 6–12 h, and after 24 h reduced to a lower level than that of untreated cells (Fig. 1C).

Another important question is how many of the androgen-regulated genes identified are directly induced or repressed by androgens. In order to address this question, detailed analyses must be done for each gene. As the first step, we will divide the androgen-regulated genes identified in this study into two groups; the cyclohexamide-sensitive group whose induction or reduction of expression by DHT is blocked by the protein synthesis inhibitor and the cyclohexamide-insensitive group whose induction or reduction is not affected by cyclohexamide. We consider that genes in the latter group are directly regulated by DHT. Preliminary results suggest that approximately 20% of the 20 genes studied by RT-PCR are cyclohexamide-insensitive. Experiments are under way to determine the cyclohexamide sensitivity of all of the 149 genes listed in Table 3. Further characterization of androgen-regulated genes may provide some clues on the transition from hormone sensitive to hormone refractory prostate cancer.

A relatively small set of genes could be analyzed at the protein level, largely due to the limited sensitivity of 2-D PAGE. Nevertheless, a substantial number of detected proteins (44 of 1031 proteins analyzed, 4.3%) were affected by DHT treatment. Some of the proteins in this

subset that were identified included the products of genes that were previously shown to be affected by androgens, namely selenium binding protein, brain specific creatine kinase, SRY protein, B23 and G3PD. Using a subtractive approach, the human selenium-binding protein gene was shown to be differentially expressed in LNCaP and reversibly down-regulated by exogenous androgen in a concentration-dependent manner, in concordance with our findings at the protein level [19]. An increase in levels of the brain specific creatine kinase B chain in response to androgen has also been described [25]. Likewise the SRY gene has been found to be responsive to androgen stimulation in LNCaP cells [18]. Also, a number of glycolytic enzymes including G3PD which were affected at the protein level have been found to be responsive to androgen stimulation [26]. Androgenic regulation of the amount and phosphorylation of the protein B23, included in our list of affected proteins, has been previously described and found to be related to the early changes associated with androgen mediated growth of the prostate [27].

Corresponding SAGE data was available for most of the proteins affected by DHT treatment that were identified. Interestingly, for most of the proteins identified, there was no appreciable concordant change at the RNA level. There are several potential explanations for this lack of concordance. A particular gene may be represented by more than one protein isoform in 2-D gels. For example, in the identified group, three proteins (myosin light chain, G3PD and selenium binding protein) were represented by more than one isoform. Thus, a source of discordance between RNA and protein data may be that the protein change is limited to a particular isoform of a protein and not to overall protein products of a particular gene. Nevertheless, a change in a particular isoform is informative and biologically meaningful and may not be predictable from RNA data. A lack of concordance between RNA and protein data may also reflect either translational control, post-translational modifications, or changes in protein turnover due to DHT treatment. Yet another explanation for a lack of concordance could be a lag time for changes at the RNA level to be reflected in a protein change. It follows from the above considerations that monitoring gene expression at both RNA and protein levels may provide complementary information that could not be ascertained by solely measuring RNA or protein.

*We would like to thank Dr. Ken Kinzler (Johns Hopkins University) for providing SAGE software. This study is partially supported by DAMD17-98-1-8501, R01CA74927 and a fund from Karmanos Cancer Institute.*

Received May 21, 2001

## 5 References

- [1] Taplin, M. E., Bubley, G. J., Shuster, T. D., Frantz, M. E., et al., *N. Engl. J. Med.* 1995, 332, 1393-1398.
- [2] Elo, J. P., Kvist, L., Leinonen, K., Isomaa, V., et al., *J. Clin. Endocrinol. Metab.* 1995, 80, 3494-3500.
- [3] Taplin, M. E., Bubley, G. J., Ko, Y. J., Small, E. J., et al., *Cancer Res.* 1999, 59, 2511-2515.
- [4] Marcelli, M., Ittmann, M., Mariani, S., Sutherland, R., et al., *Cancer Res.* 2000, 60, 944-949.
- [5] Weigel, N. L., Zhang, Y., *J. Mol. Med.* 1998, 76, 469-479.
- [6] Visakorpi, T., Hyytiäinen, E., Koivisto, P., Tanner, M., et al., *Nat. Genet.* 1995, 9, 401-406.
- [7] Koivisto, P., Kononen, J., Palmberg, C., Tammela, T., et al., *Cancer Res.* 1997, 57, 314-319.
- [8] Wallen, M. J., Linja, M., Kaartinen, K., Schleutker, J., Visakorpi, T., *J. Pathol.* 1999, 189, 559-563.
- [9] Bautista, S., Valles, H., Walker, R. L., Anzick, S., et al., *Clin. Cancer Res.* 1998, 4, 2925-2929.
- [10] Smith, C. L., Nawaz, Z., O'Malley, B. W., *Mol. Endocrinol.* 1997, 6, 657-666.
- [11] Velculescu, V. E., Zhang, L., Vogelstein, B., Kinzler, K. W., *Science* 1995, 270, 484-487.
- [12] Powell, J., *Nucleic Acid Res.* 1998, 26, 3445-3446.
- [13] Kenzelmann, M., Muhlemann, K., *Nucleic Acid Res.* 1999, 27, 917-918.
- [14] Strahler, J. R., Juick, R., Hanash, S. M., in: Creighton, T. E. [Ed.] *Protein Structure: A Practical Approach*, IRL Press, England 1989, pp. 65-92.
- [15] Gharahdaghi, F., Weinberg, C. R., Meagher, D. A., Imal, B. S., Mische, S. M., *Electrophoresis* 1999, 20, 601-605.
- [16] Liang, X., Bai, J., Liu, Y.-H., Lubman, D. M., *Anal. Chem.* 1996, 68, 1012-1018.
- [17] Hanash, S. M., in: Hames, B. D., [Ed.], *Gel Electrophoresis of Proteins*, IRL Press, Oxford, 1998, pp. 189-211.
- [18] Lau, Y. F., Zhang, J., *Mol. Carcinog.* 2000, 27, 308-321.
- [19] Yang, M., Sytkowski, A. J., *Cancer Res.* 1998, 58, 3150-3153.
- [20] Prescott, J. L., Blok, L., Tindall, D. J., *Prostate* 1998, 35, 71-80.
- [21] Swinnen, J. V., Ulrix, W., Heyns, W., Verhoeven, G., *Proc. Natl. Acad. Sci. USA* 1997, 94, 12975-12980.
- [22] Chan, P. S., Chan, L. W., Xuan, J. W., Choi, H. L., Chan, F. L., *Prostate* 1999, 41, 99-109.
- [23] Léger, J. G., Montpetit, M. L., Tenniswood, M., *Biochem. Biophys. Res. Commun.* 1987, 147, 196-203.
- [24] Miyake, H., Nelson, C., Rennie, P. S., Gleave, M. E., *Cancer Res.* 2000, 60, 170-176.
- [25] Somjen, D., Lundgren, S., Kaye, A. M., *J. Steroid Biochem. Mol. Biol.* 1997, 62, 89-96.
- [26] Tawfic, S., Goueli, S. A., Olson, M. O., Ahmed, K., *Cell. Mol. Biol. Res.* 1993, 39, 43-51.
- [27] Brooks, D. E., *Biochem. J.* 1976, 156, 527-537.

## REGULAR ARTICLE

# Web-based data warehouse on gene expression in human colorectal cancer

Emil Sagynaliev<sup>1</sup>, Ralf Steinert<sup>2</sup>, Gerd Nestler<sup>2</sup>, Hans Lippert<sup>2</sup>, Manfred Knoch<sup>1</sup> and Marc-André Reymond<sup>2</sup>

<sup>1</sup> Department of Surgery, Johanniter Krankenhaus, Stendal, Germany

<sup>2</sup> Department of Surgery, University of Magdeburg, Magdeburg, Germany

Based on biomedical literature databases, we tried a first step for constructing a gene expression "data warehouse" specific to human colorectal cancer (CRC). Results of genome-wide transcriptomic research were available from 12 studies, using various technologies, namely, SAGE, cDNA and oligonucleotide arrays, and adaptor-tagged amplification. Three studies analyzed CRC cell lines and nine studies of human samples. The total number of patients was 144. Out of 982 up- or down-regulated genes, 863 (88%) were found to be differentially expressed in a single study, 88 in two studies, 22 in three studies, 7 in four studies, and only 2 genes in six studies. Eight large-scale proteomics studies were published in CRC, using 2-D-, SDS- or free-flow electrophoresis, involving only 11 patients. Out of 408 differentially expressed proteins, 339 (83%) were found to be differentially expressed only in a single study, 16 in three studies, 10 in four studies, 3 in five, and 1 in eight studies. Confirmation at proteome level of results obtained with large-scale transcriptomics studies was possible in 25%. This proportion was higher (67%) for reproducing proteome results using transcriptomics technologies. Obviously, reproducibility and overlapping between published gene expression results at proteome and transcriptome level are low in human CRC. Thus, the development of standardized processes for collecting samples, storing, retrieving, and querying gene expression data obtained with different technologies is of central importance in translational research.

Received: December 27, 2004

Revised: March 26, 2005

Accepted: March 29, 2005

**Keywords:**

Colon neoplasms / Rectal neoplasms / Transcriptomics

## 1 Introduction

Translational research in human colorectal cancer (CRC) is applying a large spectrum of molecular biology, cellular biology, and advanced validation tools. In particular, genome-wide techniques are now applied to decipher modifications

in gene expression. In recent years, transcriptomics and proteomics tools have been broadly applied in CRC. The hope is that the new data obtained will now allow a classification of disease on molecular basis, deep insights into the pathophysiology of CRC, prognostic statements, and finally a systematic search for diagnostic and therapeutic targets.

Because of the complexity of the biological system under investigation, the most significant contribution of translational research in CRC is expected to derive not from the analysis of single experiments but from libraries of experiments. In other words, the results obtained so far by translational research tools in different clinical and experimental settings need to be compared, contrasted, and if possible synthesized. Thus, based on the biomedical literature databases, we tried a first step for constructing such "data warehouse" specific to human CRC.

**Correspondence:** Professor Marc-André Reymond, University of Magdeburg, Chairman and Head, Department of Surgery, Evangelic Hospital Bielefeld, Burgsteig 13, D-33617 Bielefeld, Germany

E-mail: marc.reymond@evkb.de

Fax: +49-521-144-4551

**Abbreviations:** cDNA, complementary DNA; CRC, colorectal cancer



The specific aims of the present study were to screen transcriptomics and proteomics studies published in CRC, and to elaborate a kind of "meta-analysis of gene expression" in CRC. Our hypothesis was that we would be able to determine a common set of genes and gene products that are up- or down-regulated in human CRC by overlapping results obtained by different authors.

## 2 Materials and methods

This is a literature survey of gene expression data published in human CRC. A Medline search of reports published in English from 1990 to 2004 using the terms "colon cancer" and "gene expression" with the limit "human" was performed and yielded 1979 articles. All abstracts were reviewed and a related article search was performed on appropriate abstracts. Articles were selected by a consensus of two reviewers (E.S., M.R.) that satisfied these predetermined criteria: sample origin (human and/or cell lines) and preparation detailed, technology for gene expression studies defined (transcriptomics and/or proteomics studies), and quantitative results on overexpression or underexpression available. Studies concerning single genes or arbitrarily selected genes were discarded. Results obtained in animal models were not considered.

Data were entered into an Excel working sheet (Microsoft, Seattle, USA). Since gene expression data were not always obtained with quantitative, but in most cases semi-quantitative gene expression analysis technologies, no expression values or ratios were entered into the database. Only genes reported as being over- or underexpressed were entered. No threshold was defined so that some genes defined as differentially expressed might have shown only

marginal differences. Only descriptive statistics are provided, which were obtained with the in-built tools of the Excel software.

Various methods were used for large-scale translational research in CRC, including transcriptomics and proteomics technologies. Results of genome-wide transcriptomic research are available from 12 studies, using various technologies, namely, SAGE (five studies), complementary DNA (cDNA) arrays (five studies), and oligonucleotide array (one study), as well as adaptor-tagged amplification (one study). Three studies analyzed CRC cell lines and nine other studies analyzed human samples. The total number of patients profiled in all studies was 144. Out of the nine studies with human samples, only two studies with a total of only 22 patients were performed using purified epithelial cells (one of them using cDNA arrays, the other one using SAGE technology). An overview of these studies is provided in Table 1.

Eight large-scale proteomics studies were published in CRC using various technologies, namely, 2-D PAGE (six studies), SDS-PAGE (one study), and free-flow electrophoresis (one study). In total, all proteome studies in CRC involved only 11 patients. An overview of these studies is provided in Table 2.

## 3 Results

One thousand two-hundred and forty genes have been reported to be dysregulated (up- and/or down-regulated) in human CRC, representing about 5% of the 20 000–25 000 human genes.

\* The complete dataset can be consulted:  
<http://www.chirurgiebethel.de>

**Table 1.** Large-scale or genome-wide transcriptomics studies in human CRC, using cell lines or human tissues. All studies together involved only 144 patients

Author	Year	Sample	Preparation	Method	Number of patients
Zhang <i>et al.</i> [18]	1997	Normal mucosa and primary tumors	Whole tissue	SAGE	2
SAGE-NET [19]	1997	Primary tumors and cell lines	Cell lines	SAGE	–
Zhang <i>et al.</i> [18]	1997	Cell lines	Cell lines	SAGE	–
Parle-McDermott <i>et al.</i> [20]	2000	Cell lines	Cell lines	SAGE	–
Yanagawa <i>et al.</i> [21]	2001	Paired normal and cancer	Whole tissue	cDNA	20
Takemasa <i>et al.</i> [22]	2001	Paired normal and primary tumors	Whole tissue	cDNA	20
Nottermann <i>et al.</i> [23]	2001	Paired normal mucosa and primary tumors, or adenoma	Whole tissue	Oligonucleotide array	22
Buckhaults <i>et al.</i> [24]	2001	Normal mucosa, adenoma and primary tumors	Purified epithelial cells	SAGE	6
Birkenkamp <i>et al.</i> [25]	2002	Paired normal and primary tumors	Whole tissue	cDNA	27
Lin <i>et al.</i> [26]	2002	Paired normal mucosa and primary tumors, or adenoma	Purified epithelial cells	cDNA	16
Muro <i>et al.</i> [27]	2003	Normal and cancer	Whole tissue	Adaptor-tagged	11
Williams <i>et al.</i> [28]	2003	Paired normal mucosa and primary tumors, normal and adenoma	Whole tissue	cDNA	20

**Table 2.** Large-scale or proteomics studies in human CRC, using cell lines or human tissues. All studies together involved only nine patients

Author	Publication year	Sample	Method	Number of patients
Reymond <i>et al.</i> [9]	1997	Purified epithelial cells	2-D PAGE	1
Simpson <i>et al.</i> [29]	2000	Cell lines	SDS-PAGE	–
Lawrie <i>et al.</i> [30]	2001	Purified epithelial cells	2-D PAGE	4
Simpson <i>et al.</i> [31]	2001	Cell lines	Free-flow electrophoresis	–
Medjahed <i>et al.</i> [32]	2003	<i>In silico</i>	2-D PAGE	–
Demalte <i>et al.</i> [33]	2003	Cell lines	2-D PAGE	–
Stierum <i>et al.</i> [34]	2003	Cell lines	2-D PAGE	–
Friedman <i>et al.</i> [35]	2004	Human whole tissue	2-D PAGE	6

The vast majority of dysregulated genes in human CRC was found using transcriptomics tools: a total of 982 genes was found to be differentially expressed in at least one of these 12 transcriptomics studies. Out of these 982 genes, 863 (88%) were found to be differentially expressed only in a single study, in other words these results have not been reproduced so far. The other findings could be reproduced in two or more transcriptomics studies: 88 genes were found to be differentially expressed in two studies, 22 in three studies, 7 in four studies, no gene in five studies, and 2 genes in six studies. The most cited genes are listed in Tables 3 (up-regulation) and 4 (down-regulation).

A total of 408 proteins were found to be differentially expressed in human CRC in at least one study. Out of these 408 molecules, 339 (83%) were mentioned in a single study, in other words these results could not be reproduced. Differential expression of the remaining 70 proteins were mentioned with following frequency: 40 were mentioned in two proteomics studies, 16 in three studies, 10 in four studies, 3 in five studies, and a single protein in eight studies. The most cited proteins are listed in Table 5. It has to be noted that only a single study [1] provided differential display protein expression data obtained in the human patient, using whole tissue biopsy.

It is also difficult to reproduce transcriptomics results with proteomics tools. Out of 982 genes found to be differentially expressed in human CRC by genome-wide transcriptomics technologies (Table 6a), only 177 (18%) have been confirmed using proteome technologies. When the genes reported to be differentially expressed in three and more transcriptomics studies ( $n = 31$ ) are compared with those reported to be differentially expressed in three and more proteomics studies ( $n = 30$ ), only two genes (actin: ACTB\_HUMAN and creatin-kinase: KCRB\_HUMAN) can be matched. Thus, the probability of reproducing a gene expression result obtained at transcriptome level is low when using proteomics technologies.

In contrast, there is a better reproducibility when proteomics results are verified using transcriptomics tools (Table 6b). In fact, when the subset of 30 proteins that are consistently (in three studies or more) reported to be dysregulated in CRC is compared with transcriptomics results, 20

genes can be matched, representing about two-thirds. Thus, the probability of being able to reproduce a proteomics result using transcriptomics tools is more than three times higher than the opposite way (67 vs. 18%). However, it has to be noted that in 16% of cases, results remain unclear or even contradictory.

Both in the transcriptomics and proteomics studies, many genes and factors were found to be differentially regulated that obviously do not play a causal role in CRC carcinogenesis.

#### 4 Discussion

This study was aimed at screening transcriptomics and proteomics studies published in CRC, and to elaborate a kind of meta-analysis of gene expression in CRC in order to generate a data warehouse that would be useful in translational research. A significant number of translational research studies have been published in CRC, starting in 1997 with SAGE technology, followed by an increasing number of array-based transcriptomics studies since 2001, and more recently by several proteomics studies. This has created a significant amount of data which we were then able to compile.

Our starting hypothesis was that we would be able to determine a common set of genes and gene products that are up- or down-regulated in human CRC by comparing results obtained by different authors. This endeavor was not very successful; obviously, overlapping between published gene expression results both at proteome and transcriptome level is low in human CRC. In fact, more than 80% of results could not be reproduced. Several explanations can be provided to explain this lack of reproducibility.

First, the number of patients included in the studies is low (altogether 144 patients in transcriptomics, 11 in proteomics studies); some high impact publications having been conducted on samples of only two patients [18]. This is a problem because interindividual genetic variability is high in human CRC [2]. Thus, it is allowed to hypothesize that some results attributed to gene dysregulation might be caused by genetic diversity rather than by cancer-specific traits.

**Table 3.** Results of transcriptomics research in human CRC. Most cited up-regulated genes out of 12 genome-wide or large-scale studies

				TRANSCRIPTOMICS											
Accession Number	Swiss-Prot	Entry Name	Name	cDNA	Adaptor-tagged	SAGE	cDNA	cDNA	cDNA	SAGE	SAGE	SAGE	SAGE	cDNA	Oligonucleotide Array
AF041260	D75363	IFM2_HUMAN	Interferon-inducible protein 1-Bd	1		1	1	1		1					
AA232508	Q9Y397	SAHH_HUMAN	S-Adenosylhomocysteine hydrolase AHcy				1	1				1	1		
AA007218	Q96QY8	BGH3_HUMAN	Transforming growth factor, beta-induced, 66 kDa	1		1	1					1			
G01119		TCPD_HUMAN	Chaperonin containing TCP-1, subunit 4 (delta)						1				1	1	
X75821	Q04984	CA11_HUMAN	COL1A1					1	1		1				
U22055	Q96AG0	TF3A_HUMAN	General transcription factor IIIA				1		1					1	
U14631	P80365	GRO_HUMAN	GRO1 oncogene (melanoma growth stimulating activity, alpha)				1						1	1	
M92439	P42704	MDP1_HUMAN	MPDP4, MDP7 microsomal dipeptidase			1	1				1				
AF029082	P31947	MPI2_HUMAN	M-phase inducer phosphatase 2	1	1	1									
M86400	P29312	PROC_HUMAN	Pyrroline 5-carboxylate reductase mRNA	1	1	1									
L76465	P15428	SPRC_HUMAN	Secreted protein, acidic, cysteine-rich (osteonectin)	1	1	1									
X57352	Q01628	Q96AG0	100 kDa coactivator mRNA	1	1										
X03205	P62888	RL30_HUMAN	50S Ribosomal protein L30	1	1										
M95787	Q01995	RL23_HUMAN	60s ribosomal protein L23	1	1										
Z25821	P42126	HS9B_HUMAN	90-kDa HSP	1	1										
U96132	Q99714	RLA2_HUMAN	Acidic ribosomal phosphoprotein P2 mRNA	1	1										
AB028893	P04643	CA14_HUMAN	Alpha-1 chain of collagen IV	1	1										
X79239	Q02546	APA1_HUMAN	Apolipoprotein A-I	1	1										
M13934	P06366	BMP4_HUMAN	Bone morphogenetic protein-2B (BMP-2B)	1	1										
M60854	P17008	CSE1_HUMAN	CAS chromosome segregation gene homologue	1	1										
M13932	P08708	CD81_HUMAN	CD81 antigen	1	1										
X69150	P25232	TCPZ_HUMAN	Chaperonin containing TCP1, subunit 6A(zeta1)	1	1										
L06498	P17075	CA21_HUMAN	Collagen alpha-2 type 1	1	1										

Table 3. Continued

				TRANSCRIPTOMICS											
Accession Number	Swiss-Prot	Entry Name	Name	cDNA	Adaptor-tagged	SAGE	cDNA	cDNA	cDNA	SAGE	SAGE	SAGE	SAGE	cDNA	Oligonucleotide Array
L04483	P35265	C02_HUMAN	Complement component C2	1	1										
M64716	P25111	R10A_HUMAN	Csa-19mRNA	1	1										
X77770	P02383	ROA1_HUMAN	DNA binding protein UPI, liver mRNA fragment	1	1										
W52460	P14798	TP2A_HUMAN	DNA topoisomerase II (top2)	1	1										
X55715	P23396	IF39_HUMAN	Eukaryotic translation initiation factor(EIF3)mRNA	1	1										
M77234	P49241	IF2B_HUMAN	Eukaryotic translation initiation factor 2, subunit 2	1	1										
M58458	P12750	GGH_HUMAN	Gamma-glytamyl hydrolase	1	1										
D16992	P10660	G3P2_HUMAN	Glyceraldehyde 3 phosphate dehydrogenase	1	1										
M77233	P23821	SYG_HUMAN	Glycyl-tRNA synthetase	1	1										
F16294	P09058	RS8_HUMAN	H19 RNA	1	1										
X61156	P08865	HS9A_HUMAN	Heat shock Protein HSP 90-Alpha	1	1										
L36055	Q13541	HMG1_HUMAN	Hmg1 mRNA for high mobility group protein I	1	1										
AA316619	P04645	RS23_HUMAN	Human homolog of yeast ribosomal protein S23 (D14530)	1	1										
L14599	Q15233	IMD2_HUMAN	IMP (inosine monophosphate) dehydrogenase 2	1	1										
X79234	P39026	IFM1_HUMAN	Interferon induced transmembrane protein 1(9-27)	1	1										
L06505	P30050	LDHB_HUMAN	Lactate dehydrogenase B(LDH-B)	1	1										
X64707	P26373	RS2_HUMAN	LLRep 3	1	1										
X56932	P40429	4F2_HUMAN	Lymphocyte activation antigen 4F2 large subunit	1	1										
L25899	P39030	NPL1_HUMAN	Nucleosome assembly protein 1-like 1	1	1										
X63527	P14118	GSHH_HUMAN	Phospholipid hydroperoxide glutathione per-oxidase	1	1										
X89401	P46778	GDFF_HUMAN	Prostate differentiation factor	1	1										
H59771	P23131	RL6_HUMAN	Ribosomal protein L 6	1	1										
M94314	P38663	RL18_HUMAN	Ribosomal protein L18(RPL18) mRNA	1	1										
L19527	P08526	RL1X_HUMAN	Ribosomal protein L18a mRNA, complete cds	1	1										
U14968	P46776	RL7_HUMAN	Ribosomal protein L7	1	1										
U14969	P46779	RL8_HUMAN	Ribosomal protein L8	1	1										
L38941	P49207	RS5_HUMAN	Ribosomal protein S5	1	1										
U12465	P42766	RS19_HUMAN	S 19 ribosomal protein	1	1										
F19234	P18077	MYC_HUMAN	V-myc myelocytomatosis viral oncogene homolog (avain)	1	1										
X66699	P12751	Q99497	EST(PARK7; DJ1)	1	1										

**Table 4.** Results of transcriptomics research in human CRC. Most cited down-regulated genes out of 12 genome-wide or large-scale studies

		TRANSCRIPTOMICS											
Entry Name	Name	cDNA	Paired normal and cancer	Yanagawa et al 2001 [21]									
		Adaptor-tagged	Normal and cancer	Muro et al 2003 [27]									
		SAGE	N.muc,adenoma and pr.tum	Buckhaults et al 2001 [24]									
		cDNA	Paired n. and pr.tumors	Birkenkamp et al 2002 [25]									
		cDNA	Paired n. and pr.tumors	Takenasa et al 2001 [22]									
		cDNA	Paired n.muc and pr.tum or adenoma	Lin et al 2002 [26]									
		SAGE	Cell lines	Parle-McDermott 2000 [20]									
		SAGE	Norm.muc and Prim.tum	Zhang et al 1997 [18]									
		SAGE	Primary tumors and cell lines	SAGE-NET et al 1997 [19]									
		SAGE	Cell line	Zhang et al 1997 [18]									
		cDNA	Paired n.muc and pr.tum,n and adenoma	Williams et al 2003 [28]									
		Oligonucleotide Array	Paired n.muc and pr.tum or adenoma	Notterman et al 2001 [23]									
CAH2_HUMAN	Carbonic anhydrase II	1		1									
CEA1_HUMAN	EST (BILIARY GLYCOPROTEIN)			1									
BENE_HUMAN	BENE	1		1									
CAH1_HUMAN	Carbonic anhydrase I	1	1										
GUAU_HUMAN	GCAP-II guanylate cyclase activator 2B; Uroguanylin; UGN	1		1									
FABL_HUMAN	L-FABP liver fatty acid-binding protein 1	1		1									
ACDS_HUMAN	Acyl-CoenzymeA dehydrogenase C-2 to C-3 short chain			1									
ADHG_HUMAN	ADH gama 2 subunit (aa-1-375)			1									
T4S3_HUMAN	Co-029; transmembrane 4 superfamily member 3		1	1									
DRA_HUMAN	Colon mucosa-associated (DRA)			1									
KCRB_HUMAN	Creatine kinase	1		1									
GUAN_HUMAN	GUCA1B guanylate cyclase activator 1B/guanylin			1									
O95784	IgG Fc-binding protein	1		1									
ITMC_HUMAN	Integral membrane protein 2C			1									
MK03_HUMAN	MAPK3			1									
FXY3_HUMAN	MAT8 protein			1									
ACTB_HUMAN	mRNA fragment encoding cytoplasmic actin	1											
SELP_HUMAN	Selenoprotein P			1									
EST1_HUMAN	Liver carboxylesterase 1 [Precursor]			1									
CFAD_HUMAN	Adipsin, complement factor DCDA (EST)			1									
ATPB_HUMAN	ATP5B ATP synthase			1									
CAHC_HUMAN	CA 12			1									
CANS_HUMAN	Calcium dependent protease (small subunit)			1									
CALM_HUMAN	Calmodulin-I (CALM1) mRNA, 3' UTR, partial sequence			1									
CAH4_HUMAN	Carbonic anhydrase IV			1									

Table 4. Continued

		TRANSCRIPTOMICS											
Entry Name	Name	cDNA	Adaptor-tagged	SAGE	cDNA	cDNA	cDNA	SAGE	SAGE	SAGE	SAGE	cDNA	Oligonucleotide Array
MT1E_HUMAN	cDNA similar to gb: M10942_cds1 human									1			1
O00748	CES 2 carboxylesterase 2				1	1							
CMGA_HUMAN	CgA				1								1
CLUS_HUMAN	clusterin												1
K1CT_HUMAN	Cytokeratin 20									1			1
KDGA_HUMAN	Diacylglycerol kinase				1								1
DTD_HUMAN	DTD sulfate transporter				1	1							
PLA8_HUMAN	EST 122594 5_									1			1
ATPA_HUMAN	F1-ATPase alpha subunit				1	1							
LEG3_HUMAN	Galectin-3 (Galactose-specific lectin 3) (MAC-2 antigen)				1	1							
ABP_HUMAN	HP-DA01 (diamine oxidase)										1		1
K1CS_HUMAN	Keratin 19						1				1		
K2C8_HUMAN	Keratin, type II cytoskeletal 8 (cytokeratin 8) (K 8) (CK 8)							1					
DHB2_HUMAN	L11708 Estradiol-17 beta- dehydrogenase 2						1						1
MT1H_HUMAN	Metallothionein 1H												1
MUC2_HUMAN	Mucin 2, intestinal/tracheal							1	1				
MEPA_HUMAN	PPH alpha gene											1	1
PA2A_HUMAN	RASF-A PLA2 gene						1					1	
MT1F_HUMAN	RNA helicase-related protein												1
SBP1_HUMAN	SBP selenium-binding protein						1	1					
MT1L_HUMAN	Serine theonine kinase 39 (STE20/SPS1 homolog,yeast)						1						1
MYL6_HUMAN	Myosin light polypeptide 6									1		1	
TETN_HUMAN	TNA tetranectin						1						1
VIPR_HUMAN	VIPR1 vasoactive intestinal polypeptide receptor 1						1						1
PLA8_HUMAN	Placenta-specific gene 8 protein											1	1
Q9NXM9	Hypothetical protein FLJ20151						1						1
T4S1_HUMAN	Transmembrane 4 superfamily member 1				1							1	
ALU2_HUMAN	Alu subfamily SB sequence contamination warning entry			1			1						
S116_HUMAN	S100 calcium-binding protein A16									1	1		
C14A_HUMAN	Dual specificity protein phosphatase CDC14A										1	1	

Second, results have been obtained using heterogeneous samples in particular cell lines, whole tissue biopsies, and epithelial cells purified from surgical specimens. Results obtained in cell lines do not allow accurate comparison between normal and cancer cells, and the presence/absence of proteins of interest has to be confirmed in biopsies. For example, when 2-D PAGE protein patterns of normal human

colonic crypts were compared with the CRC cell line LIM 1863, the proteins spots from normal crypts matched only 75–80% of the cell line spots and the relative expression levels of a large number of proteins differed [3]. When clinical biopsies are examined, results of phenotypic comparisons depend on the type of samples examined (e.g., heterogeneous whole tissue biopsies with inflammatory cells,

**Table 5.** Results of proteomics research in human CRC. Most cited up-regulated (orange), down-regulated (green), or mentioned (yellow) proteins out of 12 large-scale proteomics studies

Accession Number	Swiss-Prot	Entry Name	Name	SDS-PAGE	Free-flow electrophoresis	2-D PAGE	2-D PAGE	2-D PAGE	2-D PAGE	2-D PAGE	2-D PAGE	2-D PAGE
				Simpson et al 2000 [29]	Simpson et al 2001 [31]	Simpson et al 2003 [33]	Demalte et al 1999 [9]	Medjahed et al 2003 [32]	Lawrie et al 2001 [30]	Friedman et al 2004[35]	Sterum et al 2003[34]	
AF041260	O75363	ACTB_HUMAN	mRNA fragment encoding cytoplasmic actin	1	1	1	1	1	1	1	1	1
AA232508	Q9Y397	ACTG_HUMAN	Actin, Cytoplasmic 2 (Gamma-Actin)	1	1	1	1	1	1	1	1	1
AA007218	Q96QY8	GTP_HUMAN	Glutathione S-transferase M3 (brain)	1	1	1	1	1	1	1	1	1
X77584	G01119	THIO_HUMAN	Thioredoxin (ATL-derived factor) (ADF)	1	1	1	1	1	1	1	1	1
X75821	Q04984	ANX3_HUMAN	Annexin III (Lipocortin III)	1	1	1	1	1	1	1	1	1
U22055	Q96AG0	ANX4_HUMAN	Annexin IV (Lipocortin IV)	1	1	1	1	1	1	1	1	1
U14631	P80365	K1CR_HUMAN	Keratin, type I cytoskeletal 18 (cytokeratin 18) (K18) (CK18)	1	1	1	1	1	1	1	1	1
M92439	P42704	ATPB_HUMAN	ATP5B ATP synthase	1	1	1	1	1	1	1	1	1
AF029082	P31947	PPIA_HUMAN	Peptidyl-Prolyl Cis-Trans Isomerase A (Cyclophilin A)	1	1	1	1	1	1	1	1	1
M86400	P29312	PDA3_HUMAN	Probable protein disulfide isomerase ER-60 precursor	1	1	1	1	1	1	1	1	1
L76465	P15428	TPIS_HUMAN	Triosephosphate isomerase (TIM)	1	1	1	1	1	1	1	1	1
X57352	Q01628	TCTP_HUMAN	Translationally controlled tumor protein (TCTP)(p23)	1	1	1	1	1	1	1	1	1
X03205	P10809	CH60_HUMAN	Heat shock protein 60	1	1	1	1	1	1	1	1	1
M95787	Q01995	VINC_HUMAN	Vinculin	1	1	1	1	1	1	1	1	1
Z25821	P42126	EF1G_HUMAN	Elongation factor 1-gamma mRNA	1	1	1	1	1	1	1	1	1
U96132	Q99714	TBA1_HUMAN	Tubulin alpha-1 chain, brain specific	1	1	1	1	1	1	1	1	1
AB028893	P04643	CATD_HUMAN	Cathepsin D (EC 3.4.23.5)(GENE:CTSD)Homo sapiens	1	1	1	1	1	1	1	1	1
X79239	Q02546	EF1I_HUMAN	Elongation factor 1-alpha, mRNA	1	1	1	1	1	1	1	1	1
M13934	P06366	G3P2_HUMAN	Glyceraldehyde 3 phosphate dehydrogenase	1	1	1	1	1	1	1	1	1
M60854	P17008	GR78_HUMAN	78 kDa glucose-regulated protein precursor (GRP 78)	1	1	1	1	1	1	1	1	1
M13932	P08708	ENOA_HUMAN	Alpha enolase (2-phospho-D-glycerate hydrolyase)	1	1	1	1	1	1	1	1	1
X69150	P25232	CRTC_HUMAN	Calreticulum precursor (CR55) (calregulin)	1	1	1	1	1	1	1	1	1
L06498	P17075	COF1_HUMAN	Cofilin, non-muscle isoform	1	1	1	1	1	1	1	1	1
L04483	P35265	HS7C_HUMAN	Heat shock Cognate 71 kDa protein	1	1	1	1	1	1	1	1	1
M64716	P25111	HBB_HUMAN	Hemoglobin beta chain.(GENE: HBB) h.s	1	1	1	1	1	1	1	1	1
X77770	P02383	K2C8_HUMAN	Keratin, type II cytoskeletal 8 (cytokeratin 8) (K 8) (CK 8)	1	1	1	1	1	1	1	1	1
W52460	P14798	RM12_HUMAN	Mitochondrial 60s ribosomal protein L7/L12 precursor	1	1	1	1	1	1	1	1	1
X55715	P23396	GR75_HUMAN	Mitochondrial stress-70 protein precursor (GRP 75)	1	1	1	1	1	1	1	1	1
M77234	P49241	SODC_HUMAN	Superoxide Dismutase (Cu-Zn)	1	1	1	1	1	1	1	1	1
M58458	P12750	KCRB_HUMAN	Creatine kinase	1	1	1	1	1	1	1	1	1
D16992	P10660	143Z_HUMAN	14-3-3 protein zeta/delta (protein kinase C inhibitor protein-1)	1	1	1	1	1	1	1	1	1
M77233	P23821	RL4_HUMAN	60s ribosomal protein L4 (L1)	1	1	1	1	1	1	1	1	1
F16294	P09058	ANX1_HUMAN	Annexin I (lipocortin I) (calpactinII) (chromobindin 9) (p35)	1	1	1	1	1	1	1	1	1
X61156	P08865	ROA1_HUMAN	DNA binding protein UPI, liver mRNA fragment	1	1	1	1	1	1	1	1	1
L36055	Q13541	FABL_HUMAN	L-FABP liver fatty acid-binding protein 1	1	1	1	1	1	1	1	1	1



Table 5. Continued

Accession Number	Swiss-Prot	Entry Name	Name	SDS-PAGE	Free-flow electrophoresis	2-D PAGE	2-D PAGE	2-D PAGE	2-D PAGE	2-D PAGE	2-D PAGE	2-D PAGE
AA316619	P04645	PAB1_HUMAN	Poly(A) binding protein, mRNA	1								
L14599	Q15233	RL3_HUMAN	Ribosomal protein L 3	1								
X79234	P39026	RS19_HUMAN	S 19 ribosomal protein	1								
L06505	P30050	SAHH_HUMAN	S-Adenosylhomocysteine hydrolase AHCY)		1							
X64707	P26373	TERA_HUMAN	Transitional endoplasmic reticulum ATPase (TER ATPase)	1								
X56932	P40429	GDIR_HUMAN	Rho GDP-dissociation inhibitor 1									1
L25899	P39030	ACON_HUMAN	Aconitate hydratase, mitochondrial precursor (aconitase)	1		1						
X63527	P14118	AKA1_HUMAN	Alcohol dehydrogenase (NADP +) (aldehyde reductase)	1		1						
X89401	P46778	AAC4_HUMAN	Alpha-Actinin 4	1	1							
H59771	P23131	NPM_HUMAN	B23 nucleophosmin	1		1						
M94314	P38663	CATA_HUMAN	Catalase	1			1					
L19527	P08526	CD44_HUMAN	CD44 antigen (Phagocytic glycoprotein I)			1	1					
U14968	P46776	TCPZ_HUMAN	Chaperonin containing TCP1, subunit 6A(zeta1)	1		1						
U14969	P46779	COXA_HUMAN	Cytochrome c oxidase polypeptide Va, mitochondrial (EC 1.9.3.1) (GENE:COX5A)H.s			1	1					
L38941	P49207	COXB_HUMAN	Cytochrome C oxidase polypeptide VB precursor.COX5B	1			1					
U12465	P42766	EF1B_HUMAN	Elongation factor 1-beta (EF-1-beta)			1	1					
F19234	P18077	ER29_HUMAN	Endoplasmic reticulum protein Erp29 (Erp31) (Erp28).			1	1					
X66699	P12751	EZRI_HUMAN	Ezrin (p81) (cytovillin)(villin-2)	1		1						
D23660	P36578	ALFA_HUMAN	Fructose-Biphosphate Aldolase A (Muscle-Type Aldolase)	1	1							
U14966	P46777	DHE3_HUMAN	Glutamate dehydrogenase 1 precursor (GDH)	1			1					
X06705	P11518	HS9A_HUMAN	Heat shock Protein HSP 90-Alpha	1	1							
D14531	P32969	ROL_HUMAN	Heterogeneous nuclear ribonucleoprotein L (HNRPN L)	1		1						
U30255	P52209	K2C1_HUMAN	Keratin, type II cytoskeletal 1 (cytokeratin 1) (K1) (CK1)	1	1							
M81182	P28288	PDX1_HUMAN	Peroxisomal protein 1 (EC 1.11.1.-) (Thioredoxin peroxidase 2)			1				1		
U04627	P40939	O60506	PRM RNA binding protein Gry-rbp (GRY-RBP)mRNA	1						1		
X87949	P11021	PHB_HUMAN	Prohibitin	1		1						
M16660	P08238	PDI_HUMAN	Protein disulfide isomerase precursor (PDI)	1		1						
U79725	Q99795	SPCB_HUMAN	Spectrin beta chain, erythrocyte (Beta-I spectrin)			1	1					
D10511	P24752	PDX3_HUMAN	Thioredoxin-dependent peroxide reductase, mitochondrial			1	1					
M55040	P22303	TPM3_HUMAN	Tropomyosin alpha 3 chain (Tropomyosin 3)			1				1		
AF102542	O95395	NDKA_HUMAN	Non-metastatic cells 1, protein (NM23A)expressed in									
W45148	P24666	I43S_HUMAN	I4-3-3 PROTEIN SIGMA			1						
M17885	P05388	RL24_HUMAN	60s ribosomal protein L24(L30)	1								
M17886	P05386	APA1_HUMAN	Apolipoprotein A-I			1						
M17887	P05387	DHSA_HUMAN	SDH2 succinate dehydrogenase flavoprotein subunit	1								





Table 6. Continued

## b. Proteomics &gt; Transcriptomics

				TRANSCRIPTOMICS														PROTEOMICS																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																						
				cDNA	Paired normal and cancer	Yanagawa et al 2001 [21]	adaptor-tagged	Normal and cancer	Murn et al 2003 [27]	SAGE	N.muc,adenoma and pr.tum	Buckhaults et al 2001 [24]	cDNA	Paired n. and pr.tumors	Birkenkamp et al 2002 [25]	cDNA	Paired n. and pr.tumors	Takenasa et al 2001 [22]	cDNA	Paired n.muc and pr.tum or adenoma	Lin et al 2002 [26]	SAGE	Cell lines	Parle-McDermott 2000 [20]	SAGE	Norm.muc and Pr.tum	Zhang et al 1997 [18]	SAGE	Primary tumors and cell lines	SAGE-NET et al 1997 [19]	SAGE	Cell line	Zhang et al 1997 [18]	cDNA	Williams et al 2003 [28]	Oligonucleotide Array	Notteman et al 2001 [23]	SDS-PAGE	Simpson et al 2000 [29]	Free-flow electrophoresis	Simpson et al 2001 [31]	2D-PAGE	Denalio et al 2003 [33]	2D-PAGE	Purified epithelial cells	Reymond et al 1999 [9]	2D-PAGE	In silico	Medjahed et al 2003 [32]	2D-PAGE	Purified epithelial cells	Lawrie et al 2001 [30]	2D-PAGE	Human whole tissue	Friedman et al 2004[35]	2D-PAGE	Cell lines	Sierum et al 2003[34]																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
Accession	Number	Swiss-Prot	Entry Name	Name																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				

necrosis, blood, stool, *etc.*), so that standardized sample preparation procedures are critical for obtaining reproducible results. Several sample preparation methods have been described, in particular fluorescence-activated cell-sorting (FACS) [4], laser capture microdissection (LCM) [5–8], immunomagnetic beads separation [9], and cellular fractionation [10]. Unfortunately, these sample preparation procedures were barely applied in CRC. In our experience, the beads-based method is characterized by several advantages when compared with other cell purification procedures [11].

Third, gene expression patterns depend on the arrays technology platform. In transcriptomics studies any factors may affect the outcome of a microarray experiment, in particular technical, instrumental, computational, and interpretative factors. In fact, lack of reproducibility and accuracy is a major concern in microarray studies [12]. When cross-platform comparison was performed, reproducibility was insufficient: only four genes from a set of 185 common genes selected behaved consistently on three array platforms, and agreement of about 30% was found between two brands [13].

Fourth, in proteomics studies, 2-D PAGE or 2-D DIGE have well-known technological limitations. In CRC, even after epithelial cell enrichment using magnetic beads, the mean CV of repeated 2-D PAGE analysis with silver staining was found to lie between 20 and 28%. Only 47% (interrun) to 76% (intrarun) of spots could be matched within a triplicate experiment. Interindividual phenotypic variability was high. Thus, even under well-defined experimental conditions, 2-D PAGE parallel analysis of paired CRC samples is hampered by a significant variability [2].

Fifth, the methods applied for generating, formatting, storing, retrieving, and querying data are of outmost importance to assess methodological and biological variation in gene expression analysis. Unfortunately, due to the small sample size (number of patients), large number of variables examined at once, and absence of double or triple experiments (arrays and gels are expensive and samples are rare) statistical analysis is often not valid. In particular, assessing the reproducibility of a variable is necessary (*e.g.*, using the intraclass correlation coefficient) for comparing multiple samples at once. The use of median values instead of mean values has been shown to improve data correction [14]. It has also been proposed to use housekeeping genes as endogenous controls [15]. A dedicated society, The Microarray Gene Expression Data (MGED) Society, has been formed to facilitate the sharing of gene expression data generated by functional genomics and proteomics experiments [16].

Finally, correlation between results of transcriptomics *versus* proteomics results is low. For CRC, there is no publication comparing mRNA and protein expression for a cohort of genes. However, extrapolation is reasonable from another epithelial cancer (lung adenocarcinoma), where such comparison has been performed. Only a subset of the proteins (17%) exhibited a significant correlation with mRNA abundance [17].

Obviously, many genes and factors found to be differentially regulated (both in transcriptomics and proteomics studies) do not play a causal role in CRC carcinogenesis. For example in the studies under investigation, at least 17 mRNAs encoding ribosomal proteins were identified to be dysregulated using cDNA arrays and 39 ESTs using SAGE technology. This broad dysfunction of protein synthesis, in particular of small molecules synthesis, has been reported not only in cancer but also in several other human diseases, where etiologies have been linked to mutations in genes of the translational control machinery [36].

However, some findings might be of particular interest in human CRC. For example, a small molecule group found to be dysregulated in human CRC is the 14–3–3 proteins family. 14–3–3 proteins are ubiquitous within all eukaryotic cells and participate in protein kinase signaling pathways. In particular, they are involved in phosphorylation-dependent protein–protein interactions that control progression through the cell cycle, initiation and maintenance of DNA damage checkpoints, activation of MAP kinases, prevention of apoptosis, and coordination of integrin signaling and cytoskeletal dynamics [37]. Given the prevalence of specific 14–3–3 isoforms expression in several human epithelial cancers [38, 39], these proteins may be involved in cancer tumorigenesis and particular isoforms may be useful as therapeutic targets in human CRC.

In summary, we propose a gene expression data warehouse in human CRC that is intended to help researchers active in the field to get an overview of the data available. However, reproducibility of results obtained in different studies was disappointing.

As a matter of fact, the development of some types of unified processes for generating information, formatting, storing, retrieving, and querying data, regardless of the technology used to generate it, is of central importance for building gene expression databases. Such unified processes have been proposed (see above) but have found only limited recognition so far. This is unfortunate because these methodological issues must be solved before trying to integrate experimental data from different sources into functional proteomics studies at the bench or *in silico*. The present meta-analysis of gene expression highlights the need for including a sufficient number of patients, analyzing only purified epithelial cells, using clinical standards (such as the international classification of disease, histopathology, and staging), assessing the variability of the technology applied, considering common standards for microarray data annotation and exchange, and finally for developing software implementing these standards and promoting the sharing of these high quality, well-annotated data within the life sciences community.

In the absence of such unified processes, reproducibility of results within a laboratory will continue to be low, confirmation of results between different groups will stay difficult, and identification of diagnostic and therapeutic targets will remain a lottery. These problems are challenging the whole drug dis-

covery process from the very beginning, namely target identification, so that target validation, assay development, and prioritization of compounds remains a high risk endeavor.

*We are very grateful to G. Nestler for her excellent technical assistance.*

## 5 References

- [1] Friedman, D. B., Hill, S., Keller, J. W. et al., *Proteomics* 2004, 4, 793–811.
- [2] Ott, V., Guenther, K., Steinert, R. et al., *Pharmacogenomics J.* 2001, 1, 142–151.
- [3] Ji, H., Whitehead, R. H., Reid, G. E. et al., *Electrophoresis*, 1994, 15, 391–405.
- [4] Reymond, M. A., Sanchez, J. C., Schneider, C. et al., *Electrophoresis* 1997, 18, 622–624.
- [5] Kondo, T., Seike, M., Mori, Y. et al., *Proteomics* 2003, 3, 1758–1766.
- [6] Banks, R. E., Dunn, M. J., Forbes, M. A. et al., *Electrophoresis* 1999, 20, 689–700.
- [7] Craven, R. A., Banks, R. E., *Proteomics* 2001, 1, 1200–1204.
- [8] Craven, R. A., Totty, N., Harnden, P. et al., *Am. J. Pathol.* 2002, 160, 815–822.
- [9] Reymond, M. A., Sanchez, J. C., Hughes, G. J. et al., *Electrophoresis* 1997, 18, 2842–2848.
- [10] Szymczyk, P., Krajewska, W. M., Jakubik, J. et al., *Tumori* 1996, 82, 376–381.
- [11] Kellner, U., Steinert, R., Seibert, V. et al., *Pathol. Res. Pract.* 2004, 200, 155–163.
- [12] Shi, L., Tong, W., Goodsaid, F. et al., *Expert Rev. Mol. Diagn.* 2004, 4, 761–777.
- [13] Tan, P. K., Downey, T. J., Spitznagel, E. L. Jr. et al., *Nucleic Acids Res.* 2003, 31, 5676–5684.
- [14] Pellis, L., Franssen-van Hal, N. L., Burema, J., Keijer, J., *Physiol. Genomics* 2003, 16, 99–106.
- [15] Janssens, N., Janicot, M., Perera, T., Bakker, A., *Mol. Diagn.* 2004, 8, 107–113.
- [16] [www.mged.org/](http://www.mged.org/), consulted on December 25, 2004.
- [17] Chen, G., Gharib, T. G., Huang, C. C. et al., *Mol. Cell. Proteomics* 2002, 1, 304–313.
- [18] Zhang, L., Zhou, W., Velculescu, V. E. et al., *Science* 1997, 276, 1268–1272.
- [19] [www.sagenet.org/cancer/table1.htm](http://www.sagenet.org/cancer/table1.htm), consulted on December 1, 2003.
- [20] Parle-McDermott, A., McWilliam, P., Tighe, O., Dunican, D., Croke, D. T., *Br. J. Cancer* 2000, 83, 725–728.
- [21] Yanagawa, R., Furukawa, Y., Tsunoda, T. et al., *Neoplasia* 2001, 3, 395–401.
- [22] Takemasa, I., Higuchi, H., Yamamoto, H. et al., *Biochem. Biophys. Res. Commun.* 2001, 285, 1244–1249.
- [23] Notterman, D. A., Alon, U., Sierk, A. J., Levine, A. J., *Cancer Res.* 2001, 61, 3124–3130.
- [24] Buckhaults, P., Rago, C., St. Croix, B. et al., *Cancer Res.* 2001, 61, 6996–7001.
- [25] Birkenkamp-Demtroder, K., Christensen, L. L., Olesen, S. H. et al., *Cancer Res.* 2002, 61, 4352–4363.
- [26] Lin, Y. M., Furukawa, Y., Tsunoda, T. et al., *Oncogene* 2002, 21, 4120–4128.
- [27] Muro, S., Takemasa, I., Oba, S. et al., *Genome Biol.* 2003, 4, R21.
- [28] Williams, N. S., Gaynor, R. B., Scoggin, S. et al., *Clin. Cancer Res.* 2003, 9, 931–946.
- [29] Simpson, R. J., Connolly, L. M., Eddes, J. S. et al., *Electrophoresis* 2000, 21, 1707–1732.
- [30] Lawrie, L. C., Curran, S., McLeod, H. L., Fothergill, J. E., Murray, G. I., *Mol. Pathol.* 2001, 54, 253–258.
- [31] Simpson, R. J., *Proteomics* 2001, 1, 807–818.
- [32] Medjahed, D., Luke, B. T., Tontesh, T. S. et al., *Proteomics* 2003, 3, 1445–1453.
- [33] Demalte, I., Swiss-2-D PAGE database: Search for DLD1\_Human. <http://www.expasy.org>, Release 16 and updates up to 17-Nov-2003.
- [34] Stierum, R., Gaspari, M., Dommels, Y., et al., *Biochim. Biophys. Acta* 2003, 1650, 73–91.
- [35] Friedman, D. B., Hill, S., Keller, J. W. et al., *Proteomics* 2004, 4, 793–811.
- [36] Calkhoven, C. F., Muller, C., Leutz, A., *Trends Mol. Med.* 2002, 8, 577–583.
- [37] Wilker, E., Yaffe, M. B., *J. Mol. Cell Cardiol.* 2004, 37, 633–642 (review).
- [38] Qi, W., Liu, X., Qiao, D. et al., *Int. J. Cancer* 2005, 113, 359–363.
- [39] Moreira, J. M., Gromov, P., Celis, J. E., *Mol. Cell. Proteomics* 2004, 3, 410–419.

## Molecular profiling in prostate cancer

F. Feroze-Merzoug<sup>1</sup>, M.S. Schober<sup>1</sup> and Y.Q. Chen<sup>1,2</sup>

<sup>1</sup>Department of Pathology, <sup>2</sup>Center for Molecular Medicine and Genetics, Wayne State University, Detroit, MI, USA

**Key words:** prostate, androgen receptor, androgen regulation, differential expression, microarray, expressed sequence tags, serial analysis of gene expression, proteomics

### Abstract

Prostate cancer is the most diagnosed cancer and the second leading cause of cancer death among men in the United States. Ability to detect this cancer early and availability of better prognostic markers are critical in order to decrease morbidity and mortality of prostate cancer. With the recent development in gene expression analysis methodology, expression profiles of thousands of genes can be generated in tissue samples and cell lines. Comparison of the global gene expression patterns between normal prostate and tumors at different stages may allow us to understand better the molecular mechanism of prostate tumorigenesis and progression. Different cancer cell lines and tissues appear to have different gene expression patterns that provide a new tool to classify tumors. Molecular classification of prostate cancer holds great promise for early detection and prognosis of this disease in the future. In this review, we summarize some of the recent mRNA and protein expression profiling studies performed in prostate cancer. Further, we discuss the potential benefits and limitations of current profiling technology.

### Introduction

Prostate cancer is the most frequently diagnosed cancer and the second leading cause of cancer death among men in the United States. Considerable progress has been made in the early detection and treatment of prostate cancer over the last two decades. Nonetheless, mortality from prostate cancer remains a significant health care problem. Two critical issues for reduction of mortality and morbidity from this disease are early detection and improved prognostic markers. Interestingly, a majority of men will develop histological prostate carcinoma by their fifth decade [1], however, many will never progress to clinically relevant disease. Some clinical cancers are more aggressive, and many are slow-growing. Early detection of aggressive prostate cancer is important, especially in the organ-confined stage, as treatment at this phase will likely cure the cancer. However, even if the prostate cancer is left untreated, many men with clinically-diagnosed, less aggressive prostate cancers will not die of this disease rather other causes. With an increasing number of men with clinically-diagnosed prostate cancer, this

problem leads to an enormous cost to the health care system as well as the emotional trauma and potential physiological side effects to the individual. Therefore, development of better prognostic markers is critical to the effort to identify and treat patients with unfavorable prognosis.

Prostate is a hormone-regulated organ. Androgens affect prostatic epithelial cell proliferation and differentiation, and play a critical role in tumorigenesis. A common treatment of prostate cancer involves the deprivation of androgen, clinically known as androgen ablation. Despite the initial response of patients to hormonal therapy, the vast majority of patients eventually relapse. Molecular mechanisms involved in the development of hormone-refractory prostate cancer are unknown. However, several biological processes may contribute to this development. Mutations on the androgen receptor (AR) can cause a ligand-independent activation, or promiscuity of the receptor, i.e. alteration in the ligand binding specificity such that it binds hormones such as progesterone and estradiol [2,3]. Another phenomenon observed in 28–30% of recurrent hormone-refractory prostate cancer tumors

is AR amplification [4–6]. Increased levels of AR may allow tumor cells to survive and thrive in low androgen environments. Specific growth factors and cytokines, such as insulin-like growth factor-I (IGF-I) [7] and interleukin-6 (IL-6) [8] may also activate the AR in a ligand-independent manner. Coregulators can affect receptor transactivation depending upon the delicate balance between coactivator and corepressor expression levels [9,10]. Regardless of the mechanism leading to androgen independence, downstream genes in the androgen pathway play a critical role in the development of hormone-refractory prostate cancer.

#### Identification of differentially expressed genes

One of the keys to understanding prostate growth and tumorigenesis is the identification of differentially expressed genes between normal and tumor prostate cells, and genes regulated by androgens. Much effort has been directed to identify such genes over the years. Prostate-specific antigen (PSA) is probably the best characterized androgen-regulated gene and a common diagnostic marker. Many other genes related to growth, adhesion, cell cycle and apoptosis have been studied.

Several lines of evidence suggest that insulin-like growth factors (IGFs) may play an important role in prostate cancer. IGF-I is a potent serum marker for prostate cancer risk [11]. Increased expression of IGF-I and its receptor may be associated with tumor progression [12]. In addition, the AR can be activated by IGF-I in a ligand-independent manner [7]. A reduction in the cell adhesion molecule E-cadherin and its interacting molecule catenin may play a role in prostate cancer metastasis [13,14] and is associated with poor prognosis and survival [15,16]. Over the last decade, it has become apparent that AR requires accessory factors for optimal activation of target genes. Numerous coregulators have been identified with diverse structures and potential mechanisms of coregulation. Expression level of some AR coregulators may correlate with the grade of prostate cancer [17]. Many genes involved in cell cycle regulation such as P53, RB, P21, P27 and P16, and in apoptosis such as BCL-2 and BAX have been evaluated in prostate cancer. Recently, the tumor suppressor PTEN has also been assessed. Progresses have been made in understanding the molecular mechanism of prostate cancer development. However, critical issues remain, i.e. early detection of aggressive

tumors and markers for better prognosis of clinical disease.

#### Comprehensive gene expression analyses

With the recent development in gene expression analysis methodology, expression profiles of thousands of genes can be generated in cancer cells. Different cancer cell lines and tissues appear to have different gene expression patterns [18,19], and such differences allow for molecular classifications of cancer [20,21].

Using the well established Expressed Sequence Tag (EST) method, prostate specific cDNA libraries were constructed [22], individual cDNA clones were sequenced [22,23], and a web-based prostate expression database (PEDB) was constructed [24]. ESTs and full-length cDNA sequences derived from more than 40 human prostate cDNA libraries are maintained and represent a wide spectrum of normal and pathological conditions. Prostate ESTs were assembled into distinct species groups using the multiple alignment program CAP2 and were annotated with information from the GenBank, dbEST and UniGene databases. Differential expression of each EST species can be viewed across all libraries using a Virtual Expression Analysis Tool. The PEDB is a useful addition to the public EST project originated and organized by the National Institutes of Health (<http://www.ncbi.nlm.nih.gov/>). Prostate tumor tissues are highly heterogeneous. A given tissue often contains benign, PIN lesions and carcinoma cells of various grades as well as stromal cells. With the development of laser capture microdissection method, cells of specific types and from specific areas can be precisely dissected. Prostate cDNA libraries have also been successfully constructed from microdissected tissues [25].

Although thousands of genes can be determined, the EST method is relatively slow and expensive. Using the Serial Analysis of Gene Expression (SAGE) method, a large number of genes can be quantified more efficiently [26]. An expression profile of androgen-regulated genes from LNCaP cells has been generated by SAGE [26]. Among 123,371 transcripts analyzed, a total of 28,844 distinct SAGE tags were identified representing 16,570 genes. Some 351 genes were significantly affected by dihydrotestosterone (DHT) treatment at the RNA level ( $p < 0.05$ ), of which 147 were induced and 204 repressed by DHT treatment [27]. In another study, also conducted using LNCaP cells, a total of 83,489 SAGE tags representing 23,448 known

genes or ESTs and 1,655 potentially novel genes were identified [28]. Comparison of transcripts between control and R1881-treated LNCaP cells revealed the induction of 136 genes and repression of 215 genes in response to the synthetic androgen ( $p < 0.05$ ). These identified genes are most likely comprised of two groups; namely those regulated directly by androgen and those targeted by androgen-regulated gene products. Interestingly, both studies suggested that there were approximately 300 genes regulated by androgen and, somewhat surprisingly, there were more genes repressed than stimulated by androgen treatment in LNCaP cells. It is noteworthy that the androgen-regulated genes ( $\approx 300$  per study) identified by these two studies are different with some overlaps. Such discrepancy may be contributed to the use of DHT [27] versus a more stable synthetic androgen analog R1881 [28] and dosage of  $10^{-9}$  M [27] versus  $10^{-8}$  M [28]. Induction or repression of androgen-regulated genes is also time-dependent. Indeed, PSA was found to be induced at 4–6 h, peaked between 6 and 20 h, and gradually declined after 20 h post-treatment of DHT [27].

Differentially expressed genes between normal and prostate tumor tissues have also been determined by SAGE. In one study, a total of 133,217 transcripts were analyzed, with 35,185 distinct SAGE tags identified representing 19,287 genes. Comparison of the transcripts in normal and tumor tissue revealed 156 differentially expressed genes ( $p < 0.05$ ), of which 88 genes were up-regulated and 68 genes were down-regulated in the tumor tissue [29]. Based on SAGE data, the transcriptome for human prostate was estimated to be approximately 37,000 distinct transcripts [29]. SAGE data for another study (Library PR317 normal versus Library PR317 tumor) is available at the Cancer Genome Anatomy Project (CGAP) web site and a third independent SAGE experiment (Library PrCA-1 versus Library Normal prostate) is ongoing (<http://www.ncbi.nlm.nih.gov/SAGE/sagexpsetup.cgi>).

A comprehensive SAGE database for many human cancers, including the prostate, is accessible through the CGAP web site. The levels of gene expression can be determined virtually by searching this database using either a SAGE tag or a UniGene number. More importantly, differential expression of genes between tissues or cell lines can be explored by the xProfiler algorithm (<http://www.ncbi.nlm.nih.gov/SAGE/sagexpsetup.cgi>).

The use of DNA microarray for expression profiling has steadily gained popularity during the past five

years. Gene expression was measured in normal and organ-confined prostate cancer by 588-gene arrays [30]. Levels of the GSTM1, MCP-1, TNFR-1, TGF- $\beta$ 3 and ID-1 genes were found to be significantly reduced in tumor cells. In another study, expression profiling was performed in primary human prostate cancer and benign prostatic hyperplasia (BPH) using 6500-gene arrays. Gene expression in each of the 16 prostate cancer and 9 BPH specimens was compared with a common reference to generate normalized measures for each gene across all of the samples. Using an analysis of complete pairwise comparisons of all expression profiles, discernable patterns of overall gene expression that differentiate prostate cancer from BPH were clearly observed [31]. Further analysis of the data identified 210 genes with statistically significant differences in expression between prostate cancer and BPH [31]. Microarray analysis conducted by a different group identified IGFBP2 as one of the genes highly expressed in CWR22R hormone-refractory prostate tumor xenograft and hormone-refractory clinical tumors [32]. Protein levels of IGFBP2 were analyzed by immunohistochemistry on tissue arrays. High expression of the protein was observed in 100% of the hormone-refractory clinical tumors, and 36% of the primary tumors, but not in the benign prostatic specimens [32]. Another gene of interest, hepsin, a transmembrane serine protease, was identified by four independent studies [31,33–35]. Expression of hepsin was significantly correlated with measures of clinical outcome. Two independent SAGE analyses and a cDNA microarray study, however, did not identify hepsin as a differentially expressed gene [29,36] (<http://www.ncbi.nlm.nih.gov/SAGE/sagexpsetup.cgi>; Library PR317 normal versus Library PR317 tumor). The failure in detecting differential expression of hepsin by SAGE may be due to the difference in tissue samples.

EST, SAGE and cDNA microarray studies reveal significant and widespread differences in gene expression patterns between normal adjacent prostate, BPH, localized, metastatic and hormone-refractory prostate cancer. Gene expression analysis of prostate tissues should help to delineate the molecular mechanisms underlying prostate malignant growth and identify molecular markers for diagnostic, prognostic and therapeutic use. Indeed, many novel genes have been cloned as the result of gene expression profiling [37–42]. Gene expression patterns will undoubtedly be used for molecular classification of prostate cancer in the future.



### Proteomics

Much effort in gene expression profiling has been at the transcript level. However, it is clear that mRNA expression data alone are insufficient to predict cellular functional outcomes. For instance, mRNA expression data provide very little information about activation state, post-translational modification or localization of corresponding proteins. Moreover, there is evidence highlighting the disparity between mRNA transcript and protein expression levels. In a study of androgen-regulated genes by SAGE and proteomics, some 351 genes out of a total of 16,570 genes measured were found to be significantly affected by DHT treatment at the RNA level. Proteomic analysis demonstrated that 44 protein spots were affected at least two-fold in response to androgen, out of a total of 1,031 protein spots analyzed. The change in intensity for most of the affected proteins identified could not be predicted based on the level of their corresponding RNA [27]. It is believed that there is a tight coupling between transcription and translation in prokaryotes and disparity between the levels of mRNA and protein increases from prokaryotes to low eukaryotes in mammals. The lack of correlation between mRNA and corresponding protein is evident even in low eukaryotic cells such as yeast [43]. Therefore, it will be necessary to profile both mRNA and protein for a complete picture of how cells are altered during malignant transformation.

Compared to mRNA profiling, classical quantitative analysis of protein expression levels is much more time-consuming because proteins are analyzed one at a time. Another limitation is the fact that only proteins expressed at relatively high levels can be determined. Nevertheless, progress has been made in proteomic analysis of prostate cancer. A study, involving the collection of cells from prostate hyperplasia and prostate carcinoma, subsequently subjected to two-dimensional gel electrophoresis (2-DE) and computer assisted analysis, demonstrated considerable differences in expressed protein patterns [44]. Malignant tumors showed significant increases in the expression level of proliferating cell nuclear antigen (PCNA), calreticulin, HSP 90 and pHSP 60, oncoprotein 18(v), elongation factor 2, glutathione-S-transferase pi (GST-pi), superoxide dismutase and triose phosphate isomerase. In addition, decreases in the levels of tropomyosin-1 and -2, and cytokeratin 18 were observed in prostate carcinomas compared to prostate hyperplasias [44]. In a subsequent study, the decrease in tropomyosin

expression was confirmed in malignant tissues [45]. These studies were done with grossly dissected prostate tissues. It has been shown that 2-DE analysis can also be performed with laser capture microdissected samples [46].

The effects of androgen on protein expression have also been studied. Protein expression profiles of androgen-stimulated prostate cancer cells were generated by 2-DE. Mass spectrometric (MS) analysis of androgen-regulated proteins in these cells identified the metastasis-suppressor gene NDKA/nm23; a finding that may explain a marked reduction in metastatic potential when these cells express a functional androgen receptor pathway [47]. Using surface-enhanced laser desorption/ionization (SELDI) MS analysis, reproducible and discriminatory protein biomarker profiles can be obtained from as few as 25 cells in less than 5 min, from the time of dissection to the generation of the protein fingerprint. Furthermore, these protein pattern profiles are discriminatory for different tumor types, and reveal reproducible changes in expression as cells undergo malignant transformation. Consistent protein changes were identified in microdissected cells from patient-matched tumor and normal epithelium in three out of three malignant prostate tissue sets [48]. A reverse-phase protein microarray study, in which cell lysates from matched normal prostate epithelium, PIN and invasive prostate cancer were spotted onto slides, showed that cancer progression was associated with increased phosphorylation of Akt, a suppressor of apoptosis pathways, as well as decreased phosphorylation of ERK [49]. At the transition from histologically normal epithelium to PIN, a statistically significant surge was observed in phosphorylated Akt and a concomitant suppression of downstream apoptosis pathways which proceed the transition into invasive carcinoma [49].

### Conclusion and future directions

Functional genomics is an exciting and rapidly changing field. Array technology has significantly improved and proteomic techniques are swiftly evolving. Still, molecular profiling of prostate cancer is in its infancy. The massive amounts of data generated by mRNA and protein profiling require efficient, capable bioinformatics to extract biologically meaningful conclusions. Preliminary studies in prostate and other cancers suggest that different cancer cell lines and tissues have different gene expression patterns (see previous section)



[18,19], and such differences allow for molecular classification of cancer [20,21].

Like any emerging technology, many limitations must be overcome before it can be applied to clinical practice. Although microarrays show great promise, this methodology has not matured to the point of consistently generating robust and reliable data when used in the average laboratory [50]. It will also require the integration of other sources of biological information, such as gene identity, biochemical pathway for a given gene and pathology of tissues, to understand the cellular program of gene expression and molecular basis of cell transformation [51]. The industrialization of proteomics demands reproducible and high-throughput profiling technologies that current 2-DE cannot achieve. New technologies in protein arrays, either on chips or with self-encoded elements in solution, hold much promise for deciphering the diverse and immense proteome [52-54]. Imaging mass spectrometry is an interesting new technology that allows direct mapping of proteins present in tissue sections [55]. Protein-protein interaction mapping [56-58] and protein subcellular localization [59,60] should provide useful complementary information. In molecular profiling of cancer tissues, perhaps the most important of all, is the preparation of tissue samples. Profiling results are only as good as the tissues isolated for the experiment. Accordingly, laser capture microdissection may be the standard for tissue procurement in the future [61]. With the advances of profiling technology, expression patterning holds great promise for prognosis, early detection and molecular classification of prostate cancer.

## References

1. Sakr WA, Grignon DJ, Crissman JD, Heilbrun LK, Cassin BJ, Pontes JJ, Haas GP: High grade prostatic intraepithelial neoplasia (HGPIN) and prostatic adenocarcinoma between the ages of 20-69: An autopsy study of 249 cases. *In Vivo* 8: 439-443, 1994
2. Taplin ME, Bubley GJ, Ko YJ, Small EJ, Upton M, Rajeshkumar B, Balk SP: Selection for androgen receptor mutations in prostate cancers treated with androgen antagonist. *Cancer Res* 59: 2511-2515, 1999
3. Marcelli M, Ittmann M, Mariani S, Sutherland R, Nigam R, Murthy L, Zhao Y, DiConcini D, Puxeddu E, Esen A, Eastham J, Weigel NL, Lamb DJ: Androgen receptor mutations in prostate cancer. *Cancer Res* 60: 944-949, 2000
4. Visakorpi T, Hyytinen E, Koivisto P, Tanner M, Keinänen R, Palmberg C, Palotie A, Tammela T, Isola J, Kallioniemi OP: *In vivo* amplification of the androgen receptor gene and progression of human prostate cancer. *Nat Genet* 9: 401-406, 1995
5. Koivisto P, Kononen J, Palmberg C, Tammela T, Hyytinen E, Isola J, Trapman J, Cleutjens K, Noordzij A, Visakorpi T, Kallioniemi OP: Androgen receptor gene amplification: A possible molecular mechanism for androgen deprivation therapy failure in prostate cancer. *Cancer Res* 57: 314-319, 1997
6. Wallen MJ, Linja M, Kaartinen K, Schleutker J, Visakorpi T: Androgen receptor gene mutations in hormone-refractory prostate cancer. *J Pathol* 189: 559-563, 1999
7. Culig Z, Hobisch A, Cronauer MV, Radmayr C, Trapman J, Hittmair A, Bartsch G, Klocker H: Androgen receptor activation in prostatic tumor cell lines by insulin-like growth factor-I, keratinocyte growth factor, and epidermal growth factor. *Cancer Res* 54: 5474-5478, 1994
8. Hobisch A, Eder IE, Putz T, Horninger W, Bartsch G, Klocker H, Culig Z: Interleukin-6 regulates prostate-specific protein expression in prostate carcinoma cells by activation of the androgen receptor. *Cancer Res* 58: 4640-4645, 1998
9. Bautista S, Valles H, Walker RL, Anzick S, Zeillinger R, Meltzer P, Theillet C: In breast cancer, amplification of the steroid receptor coactivator gene AIB1 is correlated with estrogen and progesterone receptor positivity. *Clin Cancer Res* 4: 2925-2929, 1998
10. Smith CL, Nawaz Z, O'Malley BW: Coactivator and corepressor regulation of the agonist/antagonist activity of the mixed antiestrogen, 4-hydroxytamoxifen. *Mol Endocrinol* 11: 657-666, 1997
11. Chan JM, Stampfer MJ, Giovannucci B, Gann PH, Ma J, Wilkinson P, Hennekens CH, Pollak M: Plasma insulin-like growth factor-I and prostate cancer risk: A prospective study. *Science* 279: 563-566, 1998
12. Nickerson T, Chang F, Lorimer D, Smeekens SP, Sawyers CL, Pollak M: *In vivo* progression of LAPC-9 and LNCaP prostate cancer models to androgen independence is associated with increased expression of insulin-like growth factor I (IGF-I) and IGF-I receptor (IGF-IR). *Cancer Res* 61: 6276-6280, 2001
13. Morton RA, Bwing CM, Nagafuchi A, Tsukita S, Isaacs WB: Reduction of E-cadherin levels and deletion of the alpha-catenin gene in human prostate cancer cells. *Cancer Res* 53: 3585-3590, 1993
14. Umbas R, Schalken JA, Aalders TW, Carter BS, Karthaus HF, Schaafsma HB, Debruyne FM, Isaacs WB: Expression of the cellular adhesion molecule E-cadherin is reduced or absent in high-grade prostate cancer. *Cancer Res* 52: 5104-5109, 1992
15. Umbas R, Isaacs WB, Bringuier PP, Schaafsma HB, Karthaus HF, Oosterhof GO, Debruyne FM, Schalken JA: Decreased E-cadherin expression is associated with poor prognosis in patients with prostate cancer. *Cancer Res* 54: 3929-3933, 1994
16. Richmond PJ, Karayiannakis AJ, Nagafuchi A, Kaisary AV, Pignatelli M: Aberrant E-cadherin and alpha-catenin expression in prostate cancer: Correlation with patient survival. *Cancer Res* 57: 3189-3193, 1997
17. Fujimoto N, Mizokami A, Harada S, Matsumoto T: Different expression of androgen receptor coactivators in human prostate. *Urology* 58: 289-294, 2001

18. Ross DT, Scherf U, Eisen MB, Perou CM, Rees C, Spellman P, Iyer V, Jeffrey SS, Van de Rijn M, Waltham M, Pergamenschikov A, Lee JC, Lashkari D, Shalon D, Myers TG, Weinstein JN, Botstein D, Brown PO: Systematic variation in gene expression patterns in human cancer cell lines. *Nat Genet* 24: 227-235, 2000
19. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lønning PE, Borresen-Dale AL, Brown PO, Botstein D: Molecular portraits of human breast tumours. *Nature* 406: 747-752, 2000
20. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein Lønning P, Borresen-Dale AL: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* 98: 10869-10874, 2001
21. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES: Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science* 286: 531-537, 1999
22. Nelson PS, Ng WL, Schummer M, True LD, Liu AY, Bumgarner RE, Ferguson C, Dimak A, Hood L: An expressed-sequence-tag database of the human prostate: Sequence analysis of 1168 cDNA clones. *Genomics* 47: 12-25, 1998
23. Huang GM, Ng WL, Farkas J, He L, Liang HA, Gordon D, Yu J, Hood L: Prostate cancer expression profiling by cDNA sequencing analysis. *Genomics* 59: 178-186, 1999
24. Hawkins V, Doll D, Bumgarner R, Smith T, Abajian C, Hood L, Nelson PS: PEDB: The prostate expression Database. *Nucleic Acids Res* 27: 204-208, 1999
25. Krizman DB, Chuquai RF, Meltzer PS, Trent JM, Duray PH, Linehan WM, Liotta LA, Emmert-Buck MR: Construction of a representative cDNA library from prostatic intraepithelial neoplasia. *Cancer Res* 56: 5380-5383, 1996
26. Velculescu VB, Zhang L, Vogelstein B, Kinzler KW: Serial analysis of gene expression. *Science* 270: 484-487, 1995
27. Waghray A, Feroze F, Schober M, Yao F, Wood C, Puravs E, Krause M, Hanash S, Chen YQ: Identification of androgen-regulated genes in the prostate cancer cell line LNCaP by serial analysis of gene expression and proteomics analysis. *Proteomics* 1: 1327-1338, 2001
28. Xu LL, Su YP, Labiche R, Segawa T, Shanmugam N, McLeod DG, Moul JW, Srivastava S: Quantitative expression profile of androgen-regulated genes in prostate cancer cells and identification of prostate-specific genes. *Int J Cancer* 92: 322-328, 2001
29. Waghray A, Schober M, Feroze F, Yao F, Virgin J, Chen YQ: Identification of differentially expressed genes by serial analysis of gene expression in human prostate cancer. *Cancer Res* 61: 4283-4286, 2001
30. Chetcuti A, Margan S, Mann S, Russell P, Handelsman D, Rogers J, Dong Q: Identification of differentially expressed genes in organ-confined prostate cancer by gene expression array. *Prostate* 47: 132-140, 2001
31. Luo J, Duggan DJ, Chen Y, Sauvageot J, Ewing CM, Bittner ML, Trent JM, Isaacs WB: Human prostate cancer and benign prostatic hyperplasia: Molecular dissection by gene expression profiling. *Cancer Res* 61: 4683-4688, 2001
32. Bubendorf L, Kolmer M, Kononen J, Koivisto P, Mousset S, Chen Y, Mahlamaki E, Schraml P, Moch H, Willi N, Elkhouloun AG, Pretlow TG, Gasser TC, Mihatsch MJ, Sauter G, Kallioniemi OP: Hormone therapy failure in human prostate cancer: Analysis by complementary DNA and tissue microarrays. *J Natl Cancer Inst* 91: 1758-1764, 1999
33. Magee JA, Araki T, Patil S, Ehrig T, True L, Humphrey PA, Catalona WJ, Watson MA, Milbrandt J: Expression profiling reveals hepsin overexpression in prostate cancer. *Cancer Res* 61: 5692-5696, 2001
34. Dhanasekaran SM, Barrette TR, Ghosh D, Shah R, Varambally S, Kurachi K, Pienta KJ, Rubin MA, Chinnaiyan AM: Delineation of prognostic biomarkers in prostate cancer. *Nature* 412: 822-826, 2001
35. Welsh JB, Sapinoso LM, Su AI, Kern SG, Wang-Rodriguez J, Moskaluk CA, Frierson HF Jr, Hampton GM: Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer. *Cancer Res* 61: 5974-5978, 2001
36. Chaib H, Cockrell BK, Rubin MA, Macoska JA: Profiling and verification of gene expression patterns in normal and malignant human prostate tissues by cDNA microarray analysis. *Neoplasia* 3: 43-52, 2001
37. Vasmatzis G, Essand M, Brinkmann U, Lee B, Pastan I: Discovery of three genes specifically expressed in human prostate by expressed sequence tag database analysis. *Proc Natl Acad Sci USA* 95: 300-304, 1998
38. Cole KA, Chuquai RF, Katz K, Pack S, Zhuang Z, Cole CB, Lyne JC, Linehan WM, Liotta LA, Emmert-Buck MR: cDNA sequencing and analysis of POV1 (PB39): A novel gene up-regulated in prostate cancer. *Genomics* 51: 282-287, 1998
39. Nelson PS, Gan L, Ferguson C, Moss P, Gelinas R, Hood L, Wang K: Molecular cloning and characterization of prostate, an androgen-regulated serine protease with prostate-restricted expression. *Proc Natl Acad Sci USA* 96: 3114-3119, 1999
40. Xu LL, Shanmugam N, Segawa T, Sesterhenn IA, McLeod DG, Moul JW, Srivastava S: A novel androgen-regulated gene, PMEPA1, located on chromosome 20q13 exhibits high level expression in prostate. *Genomics* 66: 257-263, 2000
41. Xu J, Stolk JA, Zhang X, Silva SJ, Houghton RL, Matsumura M, Vedvick TS, Leslie KB, Badaro R, Reed SG: Identification of differentially expressed genes in human prostate cancer using subtraction and microarray. *Cancer Res* 60: 1677-1682, 2000
42. Lin B, White JT, Ferguson C, Bumgarner R, Friedman C, Trask B, Ellis W, Lange P, Hood L, Nelson PS: Part-I: A novel human prostate-specific, androgen-regulated gene that maps to chromosome 5q12. *Cancer Res* 60: 858-863, 2000

43. Gygi SP, Rochon Y, Franza BR, Aebersold R: Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* 19: 1720-1730, 1999
44. Alaiya A, Roblick U, Egevad L, Carlsson A, Franzen B, Volz D, Huwendiek S, Linder S, Auer G: Polypeptide expression in prostate hyperplasia and prostate adenocarcinoma. *Anal Cell Pathol* 21: 1-9, 2000
45. Alaiya AA, Oppermann M, Langridge J, Roblick U, Egevad L, Brindstedt S, Hellstrom M, Linder S, Bergman T, Jomvall H, Auer G: Identification of proteins in human prostate tumor material by two-dimensional gel electrophoresis and mass spectrometry. *Cell Mol Life Sci* 58: 307-311, 2001
46. Ornstein DK, Gillespie JW, Paweletz CP, Duray PH, Herring J, Vocke CD, Topalian SL, Bostwick DG, Linehan WM, Petricoin EF 3rd, Emmert-Buck MR: Proteomic analysis of laser capture microdissected human prostate cancer and *in vitro* prostate cell lines. *Electrophoresis* 21: 2235-2242, 2000
47. Nelson PS, Han D, Rochon Y, Corthals GL, Lin B, Monson A, Nguyen V, Franza BR, Plymate SR, Aebersold R, Hood L: Comprehensive analyses of prostate gene expression: Convergence of expressed sequence tag databases, transcript profiling and proteomics. *Electrophoresis* 21: 1823-1831, 2000
48. Paweletz CP, Gillespie JW, Ornstein DK, Simone NL, Brown MR, Cole KA, Wang QH, Huang J, Hu N, Yip TT, Rich WE, Kohn EC, Linehan WM, Weber T, Taylor P, Emmert-Buck MR, Liotta LA, Petricoin IEF: Rapid protein display profiling of cancer progression directly from human tissue using a protein biochip. *Drug Dev Res* 49: 34-42, 2000
49. Paweletz CP, Charboneau L, Bichsel VE, Simone NL, Chen T, Gillespie JW, Emmert-Buck MR, Roth MJ, Petricoin IE, Liotta LA: Reverse-phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. *Oncogene* 20: 1981-1989, 2001
50. Hess KR, Zhang W, Baggerly KA, Stivers DN, Coombes KR: Microarrays: Handling the deluge of data and extracting reliable information. *Trends Biotechnol* 19: 463-468, 2001
51. Noordewier MO, Warren PV: Gene expression microarrays and the integration of biological knowledge. *Trends Biotechnol* 19: 412-415, 2001
52. Zhu H, Bilgin M, Bangham R, Hall D, Casamayor A, Bertone P, Lan N, Jansen R, Bidlingmaier S, Houfek T, Mitchell T, Miller P, Dean RA, Gerstein M, Snyder M: Global analysis of protein activities using proteome chips. *Science* 293: 2101-2105, 2001
53. Haab BB, Dunham MJ, Brown PO: Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. *Genome Biol* 2: 4.1-4.13, 2001
54. Zhou H, Roy S, Schulman H, Natan MJ: Solution and chip arrays in protein profiling. *Trends Biotechnol* 19: S34-S39, 2001
55. Stoeckli M, Chaurand P, Hallahan DE, Caprioli RM: Imaging mass spectrometry: A new technology for the analysis of protein expression in mammalian tissues. *Nat Med* 7: 493-496, 2001
56. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, Qureshi-Emili A, Li Y, Godwin B, Conover D, Kalbfleisch T, Vijayadamodar G, Yang M, Johnston M, Fields S, Rothberg JMA: A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403: 623-627, 2000
57. Walhout AJ, Sordella R, Lu X, Hartley JL, Temple GE, Brasch MA, Thierry-Mieg N, Vidal M: Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* 287: 116-122, 2000
58. Stanyon CA, Finley RL Jr: Progress and potential of *Drosophila* protein interaction maps. *Pharmacogenomics* 1: 417-431, 2000
59. Simpson JC, Wellenreuther R, Poustka A, Pepperkok R, Wiemann S: Systematic subcellular localization of novel proteins identified by large-scale cDNA sequencing. *EMBO Rep* 1: 287-292, 2000
60. Remy I, Michnick SW: Visualization of biochemical networks in living cells. *Proc Natl Acad Sci USA* 98: 7678-7683, 2001
61. Craven RA, Banks RB: Laser capture microdissection and proteomics: Possibilities and limitation. *Proteomics* 1: 1200-1204, 2001

Address for offprints: Y.Q. Chen, Department of Cancer Biology, Wake Forest University, Medical Center Blvd., Winston-Salem, NC 27157; Tel: 336-7137655; Fax: 336-7137660; e-mail: yq-chen@WFUBMC.edu

62

## MOLECULAR ANALYSIS OF CANCER USING DNA AND PROTEIN MICROARRAYS

Juan Madoz-Carpide and Sam M. Hanash\*

### INTRODUCTION

Substantial progress has been made in our understanding of cancer as a multistep, complex disease that involves progressive changes in the genome and proteome. Various types of cancers share similarities as well as exhibit differences in cellular, biochemical and molecular traits. Microarray technologies have the potential of providing valuable insight regarding disease processes. The array format is now an established method for global analysis of nucleic acids, and in the past few years this approach has been adapted for protein studies (Table 1). Microarrays allow profiling of tumors' genomes, transcriptomes and proteomes at a scale unattainable previously.

Table 1. Reported and potential medical applications of DNA and protein microarrays in cancer research

DNA microarrays	Protein microarrays
Gene expression	Profiling of sera and body fluids
Mutation screening	Biomarker discovery
Genomic imbalance screening	Biochemical activities
Polymorphism genotyping	Protein-protein interactions
Diagnostics	Protein-DNA/RNA interactions
	Protein-drug interactions
	Phenotype analysis
	Epitope mapping
	Diagnostics

### CANCER PROFILING USING DNA MICROARRAYS

Genomics studies, especially profiling gene expression, using DNA array have had a tremendous impact on biomedical research, resulting in well over 1,000 published reports in 2002 alone. A substantial number of published studies dealt with cancer. Disease related

\* Department of Pediatrics, Comprehensive Cancer Center, University of Michigan, Ann Arbor 48109-094 USA

Best Available Copy

applications of DNA microarrays include uncovering unsuspected associations between genes and specific clinical features of disease that are helping devise novel molecular based disease classifications. Most published tumor studies using DNA microarrays have either examined a pathologically homogeneous set of tumors to identify clinically relevant subtypes, for example survivors vs non-survivors, or pathologically distinct subtypes belonging to the same lineage, for example limited stage vs advanced stage tumors to identify molecular correlates, or tumors of different lineages to identify molecular signatures for each lineage.

One of the landmark studies that have attracted much interest with respect to the potential contribution of DNA microarrays to uncover novel classes of tumors, is an analysis of diffuse large B-cell lymphoma, the most common subtype of non-Hodgkin's lymphoma<sup>24</sup>. Large B-cell lymphoma is a clinically heterogeneous disease. Only 40% of patients have a good response to current therapy with a prolonged survival. A systematic characterization of gene expression in this disease using DNA microarrays uncovered a diversity in gene expression that reflected variation in tumor proliferation rate, host response and differentiation state of the tumor. Two molecularly distinct forms of diffuse large B-cell lymphoma were uncovered which had gene expression patterns indicative of different stages of B-cell differentiation. One type expressed genes characteristic of germinal center B cells and had a significantly better overall survival than the second type, which expressed genes normally induced during *in vitro* activation of peripheral blood B cells. The analysis therefore identified previously undetected and clinically significant subtypes of lymphoma.

Studies to classify breast carcinomas based on gene expression profiles revealed that the tumors could be classified into a basal epithelial-like group, an ERBB2-overexpressing group and a normal breast-like group<sup>25, 26</sup>. A luminal epithelial/estrogen receptor-positive group could be divided into at least two subgroups, each with a distinctive expression profile. Survival analyses on a subcohort of patients with locally advanced breast cancer uniformly treated, in a prospective study, showed significantly different outcomes for patients belonging to the various groups, including a poor prognosis for the basal-like subtype and a significant difference in outcome for the two estrogen receptor-positive groups. In an independent study of 38 invasive breast cancers, striking molecular differences between ductal carcinoma specimens were uncovered that led to a suggested new classification for estrogen-receptor negative breast cancer<sup>27</sup>. Similarly, a study of 58 node-negative breast carcinomas discordant for ER status also uncovered a list of genes which discriminated tumors according to ER status<sup>28</sup>. Artificial neural networks could accurately predict ER status even after excluding top discriminator genes, including ER itself. Only a small proportion of the 100 most important ER discriminator genes are regulated by estradiol in MCF-7 cells.

An informative approach to analyze DNA microarray data in clinical studies is to divide such data into a training set to uncover associations between specific genes and certain clinical features of the disease, and a testing set to validate these associations. However since both training and testing sets are derived from the same pool of patients whose samples were available to the investigators, the extent to which such associations may apply to other patients not included in the study, who may have different characteristics, cannot be inferred. Beer et al. have undertaken a study of lung cancer in which the association they observed between a set of genes and patient survival was validated with a testing set of tumors they had available and further validated with an independent set of

tumors for which mic study<sup>29</sup>. Such extent uncovered between a

The numerous p technology for unco However it is substo mechanistic level osi how well RNA level imply that the predict of lung cancer. Chen allowed them to comp intensities of 165 pto 76 lung adenocarcino samples, mRNA level genes (21.4%) had a  $t > 0.2445$ ;  $P < 0.05$ ; isoforms of the same regulation of protein a

#### PROTEIN MICROARR

Despite the advan gaps remain both in effective strategies for investigating diseases approaches. DNA mi for uncovering directl in proteins that are not

Unlike DNA mic levels, there is a nee; different features of p and determination of various small ligands obtained using cDN, utilized for different t tissue-derived sample: total protein lysates microarrays were pres detecting array. With support at a defined l molecular interaction. ligands used to profile The whole process of nature of the support, the particle being att assays of this magniti

associations between novel molecular based microarrays have either clinically relevant or distinct subtypes used stage tumors to identify molecular

a with respect to the types of tumors, is an type of non-Hodgkin's disease. Only 40% of survival. A systematic microarrays uncovered a robustness rate, host distinct forms of diffuse patterns indicative of ones characteristic of I than the second type, of peripheral blood B I clinically significant

offices revealed that the ERBB2-overexpressing oregon receptor-positive distinctive expression advanced breast cancer different outcomes for asia for the basal-like oregon receptor-positive a, striking molecular has led to a suggested similarly, a study of 58 covered a list of genes neural networks could r genes, including ER lscriminial genes are

ical studies is to divide afic genes and certain associations. However cool of patients whose each associations may lfferent characteristics, ; cancer in which the d was validated with a an independent set of

tumors for which microarray data was collected by another group not associated with this study<sup>37</sup>. Such extensive validation clearly indicated the robustness of the associations uncovered between a set of genes and survival in lung adenocarcinoma.

The numerous published studies using DNA microarrays justify the use of this technology for uncovering patterns of gene expression that are clinically informative. However it is substantially more difficult to develop an understanding of disease at a mechanistic level using DNA microarrays. For most of the published studies it is unclear how well RNA levels reported correlate with protein levels. A lack of correlation may imply that the predictive property of the gene(s) is independent of gene function. In studies of lung cancer, Chen et al. collected both DNA microarray and 2-D PAGE data, which allowed them to compare mRNA and protein levels in the same tumors<sup>38</sup>. The integrated intensities of 165 protein spots representing protein products of 98 genes were analyzed in 76 lung adenocarcinomas and 9 unaffected lung tissues using 2-D gels. For the same 85 samples, mRNA levels were determined using oligonucleotide microarrays. Only 21 of 98 genes (21.4%) had a statistically significant correlation between protein and mRNA levels ( $r > 0.2445$ ;  $P < 0.05$ ). The mRNA/protein correlation coefficients also varied between isoforms of the same protein, indicating potentially isoform-specific mechanisms for the regulation of protein abundance.

#### PROTEIN MICROARRAYS IN CANCER RESEARCH

Despite the advances in our understanding of the molecular basis of cancer, substantial gaps remain both in our understanding of cancer pathogenesis and in the development of effective strategies for early diagnosis and for treatment. A proteomic approach to investigating diseases such as cancer may overcome some of the limitations of other approaches. DNA microarrays have limited utility for the analysis of biological fluids and for uncovering directly in the fluid, assayable biomarkers. Numerous alterations may occur in proteins that are not reflected in changes at the RNA level.

Unlike DNA microarrays that provide one measure of gene expression, namely RNA levels, there is a need to implement protein microarray strategies that address the many different features of proteins including determination of their levels in biological samples, and determination of their selective interactions with other proteins, antibodies, drugs or various small ligands. Arrays that incorporate antibodies<sup>39-41</sup> or recombinant proteins obtained using cDNA expression libraries<sup>42-43</sup> or phage-display libraries<sup>44</sup> have been utilized for different types of protein based assays. With other types of microarrays, whole tissue-derived samples have been directly arrayed onto slides, to assess the reactivity of total protein lysates with specific ligand<sup>44-46</sup>. Two practical applications of protein microarrays were presented by Kodadek<sup>40</sup>, designated protein function array and protein-detecting array. With protein function arrays, a large amount of protein is spotted on a solid support at a defined location and tested to characterize either a biochemical activity or a molecular interaction. The protein-detecting array consists of an arrayed set of protein ligands used to profile gene expression and draw signatures indicative of the cellular state. The whole process of assembling the protein array requires considerations related to the nature of the support, the type of immobilization, as well as the molecular architecture of the particle being attached. Four main different supports have been optimized to perform assays of this magnitude: chemically modified glass slides (poly-L-lysines, polyaldehydes,

boronic acid derivatives, chelates to poly-Histidines, etc.), nitrocellulose membranes or polyacrylamide gel on glass slides, and microwell plates. Each type of support exhibits advantages and disadvantages, as recently noted by Zhu and Snyder<sup>46</sup>.

One of the limitations of most of the current protein microarray technologies is the lack of control over orientation in the immobilization process. It has been shown repeatedly that optimization of physical interactions between immobilized macromolecules, e.g., antigens, antibodies, peptides, and their corresponding target ligands affects sensitivity<sup>44,45</sup>. There is currently a substantial variety of procedures for oriented immobilization: ionic interaction, specific covalent binding, apoenzyme reconstituted on the surface that binds to a prosthetic group, receptor/ligand interactions, specific affinity motifs engineered into the surface of proteins, etc. In most cases it has been shown that optimal binding of protein to solid supports requires hydrophilic spacers<sup>46</sup>. Further optimization of the arraying approach should include coupling of protein separation technologies with techniques for orientational control, that would permit different surface orientations of a given protein to interact with other proteins or ligands, to enhance efficiency and reproducibility.

The compelling need for protein chips has led numerous biotechnology companies to devise novel strategies for producing biochips for various applications. New classes of capture agents include aptamers (SomaLogic, <http://www.somallogic.com>), ribozymes (Archimix, <http://www.archimix.com>), partial-molecule imprints (Aspira Biosystems, <http://www.aspirabio.com>) and modified binding proteins (Phyllos, <http://www.phyllos.com>). For assays of protein interaction, biochips that contain either peptides or proteins are being produced. Peptides may be synthesized in very large numbers directly on the chip<sup>47</sup> (Figure 1). Alternatively, recombinant proteins may be arrayed and effort is underway to assemble large sets of purified recombinant proteins for microarrays and other applications. As an example of innovative approaches in protein biochips, a bioanalytical system based on a planar waveguide technology has been developed which allows multiplexed, quantitative biomolecular interaction analysis to be performed with high sensitivity in a microarray format. The analytical system comprises microarray chips with integrated microfluidics and a highly sensitive fluorescence reader<sup>48</sup>. Applications of such a system include both protein expression profiling and studies of protein-protein interactions. Important requirements for protein biochips include ability of the capture agents to bind their ligands linearly across the entire set of capture agents deposited or synthesized on the chip, and with adequate sensitivity and dynamic range.

There is intense interest in applying proteomics to disease marker identification. Approaches to that effect include comparative analysis of protein expression in normal and cancer tissues to identify aberrantly expressed proteins that may represent novel markers, analysis of secreted proteins in cell lines and primary cultures and direct serum protein profiling to uncover potential markers. There has been recently substantial interest in the potential of mass spectrometry to yield comprehensive profiles of peptides and proteins in biological fluids without the need to first carry out protein separations. In principle, such an approach is highly suited for marker identification because of reduced sample requirements and high throughput. This approach is currently popularized, particularly for serum analysis, by the technology referred to as surface-enhanced laser-desorption/ionization (SELDI)<sup>49</sup>. Microliter quantities of serum from many samples are applied to the surface of a protein-binding plate, with properties to bind a class of proteins. The bound proteins are treated and analyzed by matrix-assisted laser-desorption/ionization (MALDI). The mass spectra patterns obtained for different samples reflect the protein and peptide contents of

these samples. Patterns of remarkable accuracy, have protein arrays with mass as a tool with which to profile (Ciphergen Inc. USA), and, more recently, had involvement of proteins fingerprints from which analysis of tissues or biofluids with a lower molecular mass than the masses observed. For direct mass spectrometric



Figure 1. Representative capillary electrophoresis (CE) chromatograms.

Cancer tissue profiling emerge. As a model to better microenvironment, Kneze squamous cell carcinomas high-throughput proteomic total protein from specific quantitative, and potential proteins within epithelial. Differential expression of adjacent to regions of duct the epithelium. Most of the transduction pathways.

whose membranes or  
of support exhibits

technologies is the lack  
shown repeatedly that  
ecules, e.g., antigens,  
activity<sup>64,65</sup>. There is  
on: ionic interaction,  
binds to a prosthetic  
d into the surface of  
of protein to solid  
arraying approach  
ques for orientational  
stein to interact with

ology companies to  
ms. New classes of  
ic.com/), ribozymes  
'Aspirin Biosystems,  
roteins (Phytoc,  
that contain either  
very large numbers  
may be arrayed and  
elus for microarrays  
protein biochips, a  
on developed which  
be performed with  
as microarray chips  
<sup>66</sup>. Applications of  
of protein-protein  
ility of the capture  
agents deposited or  
c.

ster identification.  
sion in normal and  
sent novel markers.  
irect serum protein  
ntial interest in the  
ides and proteins in  
n principle, such an  
ample requirements  
icularly for serum  
orption ionization  
l to the surface of a  
bound proteins are  
MALDI). The mass  
peptide contents of

these samples. Patterns that distinguish between cancer patients and normal subjects with remarkable accuracy, have been reported for several types of cancer<sup>67</sup>. The coupling of protein arrays with mass spectrometry technologies is likely to become a powerful analytic tool with which to profile protein expression. Such an approach, known as ProteinChip (Ciphergen Inc, USA), was successfully applied to study prostate and ovarian cancers<sup>68</sup> and, more recently, head and neck cancers<sup>69</sup>. Results from these studies revealed the involvement of proteins in carcinogenesis processes and specifically identified protein fingerprints from which cancer biomarkers were extracted. The major drawbacks of direct analysis of tissues or biological fluids by MALDI is the preferential detection of proteins with a lower molecular mass and the difficulty in identifying the proteins corresponding to the masses observed. Further technological improvements could enhance the utility of direct mass spectrometric analysis of tissues and biological fluids.

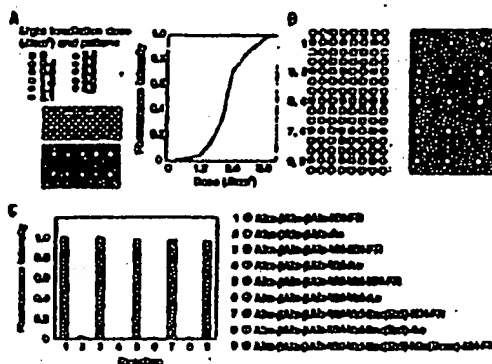


Figure 1. Representative experiments for optimization of POA deprotection. (Reproduced from Petheis *et al.*, 2002).

Cancer-tissue profiling studies that have utilized protein microarrays are beginning to emerge. As a model to better understand how patterns of protein expression shape the tissue microenvironment, Knezevic *et al.* analyzed protein expression in tissue derived from squamous cell carcinomas of the oral cavity through an antibody microarray approach for high-throughput proteomic analysis<sup>69</sup>. Utilizing laser capture microdissection to procure total protein from specific microscopic cellular populations, they demonstrated that quantitative, and potentially qualitative, differences in expression patterns of multiple proteins within epithelial cells reproducibly correlated with oral cavity tumor progression. Differential expression of multiple proteins was found in stromal cells surrounding and adjacent to regions of diseased epithelium that directly correlated with tumor progression of the epithelium. Most of the proteins identified in both cell types were involved in signal transduction pathways. They hypothesized therefore that extensive molecular



communications involving complex cellular signaling between epithelium and stroma play a key role in driving oral cavity cancer progression.

A clinically relevant application of protein microarrays is the identification of proteins that induce an antibody response in autoimmune disorders<sup>40</sup>. Microarrays were produced by attaching several hundred proteins and peptides to the surface of derivatized glass slides. Arrays were incubated with patient serum, and fluorescent labels were used to detect autoantibody binding to specific proteins in autoimmune diseases, including systemic lupus erythematosus and rheumatoid arthritis. Such microarrays represent a powerful tool to study immune responses, in a variety of diseases including cancer.

A reverse phase protein array approach that immobilizes the whole repertoire of a tissue's proteins has been developed<sup>41</sup>. A high degree of sensitivity, precision and linearity was achieved, making it possible to quantify the phosphorylated status of signal proteins in human tissue cell subpopulations. Using this approach Pawelczak et al.<sup>42</sup> have longitudinally analyzed the state of pro-survival checkpoint proteins at the microscopic transition stage from patient matched histologically normal prostate epithelium to prostate intraepithelial neoplasia and to invasive prostate cancer. Cancer progression was associated with increased phosphorylation of Akt, suppression of apoptosis pathways, as well as decreased phosphorylation of ERK. At the transition from histologically normal epithelium to intraepithelial neoplasia, a statistically significant surge in phosphorylated Akt and a concomitant suppression of downstream apoptosis pathways preceding the transition into invasive carcinoma were observed.

A major challenge in making biochips for global analysis of protein expression is the current lack of comprehensive sets of genome scale capture agents such as antibodies. As a result, biochips that target specific classes of proteins such as kinases or cytokines are much easier to produce, that would have clinical utility. Another important consideration in protein microarrays is that proteins undergo numerous post-translational modifications eg phosphorylations, glycosylations, which are highly important to their functions, as they can determine activity, stability, localization and turnover. To address the need for comprehensive analysis of proteins in their modified forms, several approaches to the liquid based separation of cell and tissue lysates were investigated in order to obtain protein fractions with reduced complexity or pure individual proteins<sup>43</sup>. The separation products can be arrayed in a manner that allows the probing of protein constituents of cells and tissues to uncover specific targets. For example, using a combination of anion exchange and reverse phase LC, Madaz-Curpido et al. have obtained some 2000 individual protein fractions that have been utilized to produce microarrays that interrogate cancer cell proteomes. Fractions that react with specific probes are within the reach of chromatographic and gel based separation techniques for resolving their individual protein constituents and of mass spectrometric techniques for identification of their constituent proteins. The LC procedures allow sufficient protein amounts to be resolved for the construction of large numbers of microarrays from a given cell or tissue source.

Protein microarrays of different types are likely to become commercially available for assays of broad sets of proteins and may well rival or at least complement DNA microarrays as tools for global expression analysis.

## SUMMARY

In conclusion, arrays of proteins on solid supports provide a powerful tool to study the vast amount of information encoded in the genome and proteome and to develop individualized therapy for

## References

1. Perou, C.M., Alizadeh, A.A., Eisen, M.B., et al. Molecular classification of human breast tumors. *Nature* 406, 746-752 (2003).
2. Alizadeh, A.A., Miller, R., Davis, S., et al. Deregulation of microRNAs in human breast cancer. *Nature* 426, 85-90 (2003).
3. Perou, C.M., Alizadeh, A.A., Eisen, M.B., et al. Molecular classification of human breast tumors. *Nature* 406, 746-752 (2003).
4. Sorlie, T., Perou, C.M., Pao, A., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
5. Brent, D., Eisen, M.B., Alizadeh, A.A., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
6. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
7. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
8. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
9. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
10. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
11. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
12. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
13. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
14. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
15. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
16. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
17. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
18. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
19. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
20. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
21. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
22. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
23. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
24. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
25. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
26. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
27. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
28. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
29. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
30. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
31. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
32. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
33. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
34. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
35. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
36. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
37. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
38. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
39. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
40. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
41. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
42. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).
43. Chen, G., Chen, G., Chen, G., et al. Gene expression patterns of human breast carcinomas confirmed by cluster analysis and protein expression data. *Proc Natl Acad Sci USA* 100, 8329-8334 (2003).

telium and stroma play

entification of proteins  
arrays were produced  
terialized glass slides.  
s were used to detect  
cluding systemic lupus  
ni a powerful tool to

whole repertoire of a  
precision and linearity  
as of signal proteins in  
l.<sup>44</sup> have longitudinally  
scope transition stage  
prostate intraepithelial  
ociated with increased  
a well as decreased  
normal epithelium to  
thorylated Akt and a  
ing the transition into

tein expression is the  
ch as antibodies. As a  
or cytokines are much  
stant consideration in  
onal modifications eg  
functions, as they can  
ldress the need for  
approaches to the liquid  
rder to obtain protein  
separation products  
ndments of cells and  
on of anion exchange  
100 individual protein  
terrogate cancer cell  
ithin the reach of  
sets individual protein  
n of their constituent  
be resolved for the  
is source.  
mercially available for  
st complement DNA

## SUMMARY

In conclusion, array-based technologies have emerged that contribute to profiling issues at the genomic, transcriptomic and proteomic levels. Analytical tools are needed to mine the vast amount of data generated. Ultimately the molecular analysis of cancer at a genome and proteome scale will allow better classification of disease and tailored individualized therapy for individual patients.

## References

1. Petricoin, E.F., Zoon, K.C., Kohn, E.C., Barnea, J.C. & Limm, L.A. Clinical proteomics: translating benchside promise into bedside reality. *Nature Reviews Drug Discovery* 1, 687-693 (2002).
24. Alizadeh, A.A. et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403, 503-511 (2000).
25. Perou, C.M. et al. Molecular portraits of human breast tumours. *Nature* 406, 747-752 (2000).
26. Sorlie, T. et al. Gene expression patterns of breast carcinomas distinguish tumor subtypes with clinical implications. *Proceedings of the National Academy of Sciences, USA* 98, 10869-10874 (2001).
27. Brannan, J.D., Aparicio, S.A. & Caldas, C. Molecular profiling of breast cancer: potential but not phylogeny. *Breast Cancer Research* 3, 71-80 (2001).
28. Gruvberger, S. et al. Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. *Cancer Research* 61, 5979-5984 (2001).
29. Beer, D.G. et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nature Medicine* 8, 816-824 (2002).
30. Chen, Q. et al. Proteomic analysis of lung adenocarcinoma: identification of a highly expressed set of proteins in cancer. *Clinical Cancer Research* 8, 2290-2305 (2002).
31. Microarray standards at last. *Nature* 419, 323 (2002).
32. Gu, H. UPA, a universal protein array system for quantitative detection of protein-protein, protein-DNA, protein-RNA and protein-ligand interactions. *Nucleic Acids Research* 28, e3 (2000).
33. Rowe, C.A., Scruggs, S.B., Feldstein, M.J., Golden, J.P. & Light, P.S. An array immunosensor for simultaneous detection of clinical analytes. *Analytical Biochemistry* 71, 433-439 (1999).
34. Mendoza, L.O. et al. High-throughput microarray-based enzyme-linked immunosorbent assay (ELISA). *Biochemicals* 27, 776-780, 782-785, 789 (1999).
35. Sillet, J.W., Curock, B., Dudson, C., Tsay, T. & Obrowski, R.J. Mass-screening, multi-analyte microarray immunoassay with imaging detection. *Clinical Chemistry* 44, 2006-2043 (1998).
36. Aronov, P. et al. Protein microchips: use for immunoassay and enzymatic reactions. *Analytical Biochemistry* 278, 123-131 (2000).
37. Hach, B.B., Dunham, M.J. & Brown, P.O. Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. *Genome Biology* 2, Research0004.1-0004.13 (2001).
38. MacBeath, G. & Schreiber, S.L. Printing proteins as microarrays for high-throughput function determination. *Science* 289, 1760-1763 (2000).
39. Schweitzer, B. et al. Immunoarrays with rolling circle DNA amplification: a versatile platform for ultrasensitive antigen detection. *Proceedings of the National Academy of Sciences, USA* 97, 10113-10119 (2000).
40. Buxnow, K. et al. A method for global protein expression and antibody screening on high-density filters of an arrayed cDNA library. *Nucleic Acids Research* 26, 5007-5008 (1998).
41. Lusk, A. et al. Protein microarrays for gene expression and antibody screening. *Analytical Biochemistry* 270, 103-111 (1999).
42. Zhu, H. et al. Analysis of yeast protein kinases using protein chips. *Nature Genetics* 26, 283-289 (2000).
43. DeWald, R.M.T., Mundy, C.R., Corrick, B.D. & Tomlinson, L.M. Antibody arrays for high-throughput screening of antibody-antigen interactions. *Nature Biotechnology* 18, 989-994 (2001).
44. Konecna, J. et al. Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nature Medicine* 4, 844-847 (1998).

45. Kallioniemi, O.P., Wagner, U., Kinnunen, J. & Seuter, G. Tissue microarray technology for high-throughput molecular profiling of cancer. *Human Molecular Genetics* 10, 657-662 (2001).
46. Pavlakis, C.P. et al. Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. *Oncogene* 20, 1981-1989 (2001).
47. Pellois, J.P. et al. Individually addressable parallel peptide synthesis on microchips. *Nature Biotechnology* 20, 912-916 (2002).
48. Pawlak, M. et al. Zepioteins' protein microarrays: a novel high performance microarray platform for low abundance protein analysis. *Proteomics* 2, 383-393 (2002).
49. Kaczevic, V. et al. Proteomic profiling of the cancer microenvironment by antibody arrays. *Proteomics* 1, 1271-1278 (2001).
50. Robinson, W.H. et al. Autoantigen microarrays for multiplex characterization of autoantibody responses. *Nature Medicine* 8, 293-301 (2002).
51. Madraz-Gurpida, J., Wang, H., Muzek, D.E., Brichory, F. & Hamash, S.M. Protein based microarrays: A tool for probing the proteomes of cancer cells and tissues. *Proteomics* 1, 1279-1287 (2001).
52. Yamamoto, A. et al. Infrequent presence of anti-c-Myc antibodies and absence of c-Myc oncoprotein in sera from lung cancer patients. *Oncology* 56, 129-133 (1999).
53. Stockert, E. et al. A survey of the humoral immune response of cancer patients to a panel of human tumor antigens. *Journal of Experimental Medicine* 187, 1349-1354 (1998).
54. Gourevitch, M.M. et al. Polymorphic epithelial mucin (MUC-1)-containing circulating immune complexes in carcinoma patients. *British Journal of Cancer* 72, 934-938 (1995).
55. Gure, A.O. et al. Human lung cancer cell lines recognized by autoantigenic antibodies: Definition of a novel cDNA derived from the tumor suppressor gene locus on chromosome 3p21.3. *Cancer Research* 58, 1004-1011 (1998).
56. Yamamoto, A., Shimizu, E., Ogura, T. & Sone, S. Detection of anti-antibodies against L-myc oncogene products in sera from lung cancer patients. *International Journal of Cancer* 23, 283-289 (1995).
57. Soussi, T. The humoral response to the tumor-suppressor gene product p53 in human cancer: implications for diagnosis and therapy. *Immunology Today* 17, 354-356 (1996).
58. Old, L.J. & Chen, Y.T. New paths in human cancer serology. *Journal of Experimental Medicine* 187, 1163-1167 (1998).
59. Le Naou, P. et al. A distinct repertoire of autoantibodies in hepatocellular carcinoma identified by proteomic analysis. *Molecular and Cellular Proteomics* 1, 197-203 (2002).
60. Kodrzyk, T. Protein microarrays: Prospects and problems. *Chemical Biology* 8, 103-115 (2001).
61. Zhu, H. et al. Protein arrays and microarrays. *Current Opinion in Chemical Biology* 5, 40-45 (2001).
62. Senior, R. Fingerprinting disease with protein chip arrays. *Molecular Medicine Today* 5, 326-327 (1999).
63. Eggeling, F. et al. Tissue-specific microdissection coupled with ProteinChip array technologies. Applications in cancer research. *Bioarchitectures* 29, 1066-1070 (2000).
64. Madraz-Gurpida, J. et al. Modification of electrocatalytic NADPH oxidation through oriented immobilization of ferredoxin: NADP(+) reductase onto modified gold electrodes. *Journal of the American Chemical Society* 122, 9808-9817 (2000).
65. Delamar, B. et al. Immobilization of antibodies on a photoactive self-assembled monolayer on gold. *Langmuir*, 12, 1997-2006 (1996).
66. Muller, W. et al. Attempts to mimic docking processes of the immune system: recognition-induced formation of protein multilayers. *Science* 262, 1704-1708 (1993).

## PROTEOMIC AI

Julie D. Walfkuhle<sup>1,\*</sup>, C  
and Lehee Liotta<sup>1</sup>

<sup>1</sup>FDA/NCI Clinical Pro  
Research, National Can  
Laboratory of Pathology,  
MD 20892. <sup>2</sup>FDA/NCI  
Research, Food and Drug

## ABSTRACT

The field of proteomic detection and diagnosis of efficacy and toxicity. A goal. Two-dimensional discovery-based proteomic (LCM) and highly sensitive greater numbers of low defined cell populations enable the identification Surface-enhanced laser of the high throughput chart may be best suited for di to supplement our arsen cellular networks and not and that we will soon b before, during and after ti

\* To whom correspondence sh  
535, 8800 Rockville Pike, Be  
† Current address: Departme  
Uniformed Services School of

New Trends in Cancer for the 21  
Llambert-Bouch and Feilpo, Kto

## Genome-Wide Profiling of Gene Expression in 29 Normal Human Tissues with a cDNA Microarray

Akihiko SAITO-HISAMINATO,<sup>1,†</sup> Toyomasa KATAGIRI,<sup>1,†</sup> Soji KAKIUCHI,<sup>1</sup> Tohru NAKAMURA,<sup>1</sup> Tatsuhiko TSUNODA,<sup>2</sup> and Yusuke NAKAMURA<sup>1,\*</sup>

*Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1, Shirokanedai, Minato-ku, Tokyo 108-8639, Japan<sup>1</sup> and Laboratory for Medical Informatics, SNP Research Center, Riken (Institute of Physical and Chemical Research), 4-6-1, Shirokanedai, Minato-ku, Tokyo 108-8639, Japan<sup>2</sup>*

(Received 14 February 2002; revised 25 March 2002)

### Abstract

We have performed a comprehensive analysis of the expression profiles in 25 adult and 4 fetal human tissues by means of a cDNA microarray consisting of 23,040 human genes. This study revealed a number of genes that were expressed specifically in each of those tissues. Among the 29 tissues examined, 4,080 genes were highly expressed (at least a five-fold expression ratio) in one or only a few tissues and 1,163 of those were expressed exclusively (more than a ten-fold higher expression ratio) in a particular tissue. Expression of some of the genes in the latter category was confirmed by northern analysis. A hierarchical clustering analysis of gene-expression profiles in nerve tissues (adult brain, fetal brain, and spinal cord), lymphoid tissues (bone marrow, thymus, spleen, and lymph node), muscle tissues (heart and skeletal muscle), or adipose tissues (mesenteric adipose and mammary gland) identified a set of genes that were commonly expressed among related tissues. These data should provide useful information for medical research, especially for efforts to identify tissue-specific molecules as potential targets of novel drugs to treat human diseases.

**Key words:** expression profile; normal human tissues; cDNA microarray; tissue-specific expression

### 1. Introduction

The Human Genome Project and corporate efforts achieved a complete draft sequence of the genome recently.<sup>1,2</sup> Whole-genome sequencing and large-scale expressed sequence tag (EST) projects have revealed that the human genome contains 30,000–40,000 genes.<sup>1–3</sup> In the post-sequencing era, investigating the biological and physiological functions of gene products, especially those related to diseases, is becoming increasingly important. To address this task, it is useful to know the expression properties of each gene in human tissues and organs.

cDNA microarray technology has developed dramatically over the past 5 years, and we can now obtain expression profiles of thousands of genes simultaneously. Microarrays have been used to investigate gene-expression profiles in yeast,<sup>4</sup> *Mycobacterium*,<sup>5</sup> cell lines<sup>6</sup> and many types of human cancers.<sup>7–9</sup> Moreover, statistical methods

such as cluster analysis have revealed molecular classifications based on differential expression of certain genes.

Some genes that are expressed in a tissue-specific manner have been screened using a differential cDNA display technique,<sup>10–12</sup> the "Body Map" system,<sup>13–16</sup> or by computational subtraction in serial analyses of gene expression (SAGE).<sup>17</sup> Obtaining information about genes expressed only in a single tissue is a critical step toward understanding the mechanism of tissue-specific transcriptional regulation or for investigating genes responsible for hereditary syndromes when the phenotype is specific to a particular organ. With that goal in mind, we have performed comprehensive expression profiling in 25 adult and 4 fetal human tissues using a cDNA microarray containing 23,040 genes, and evaluated the power of this analysis for identifying tissue-specific molecules.

Communicated by Michio Oishi

\* To whom correspondence should be addressed. Tel. +81-3-5449-5372, Fax. +81-3-5449-5433, E-mail: yusuke@ims.u-tokyo.ac.jp

† These authors contributed equally to this work.

## 2. Materials and Methods

### 2.1. Preparation of samples and T7-based amplification of RNA

Poly(A)<sup>+</sup> RNAs isolated from bone marrow, brain, heart, kidney, liver, lung, lymph node, mammary gland, pancreas, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thymus, thyroid, trachea, uterus, fetal brain, fetal kidney, fetal liver, fetal lung (Clontech), colon, and ovary (Biochain) were used as probes for the cDNA microarray. From normal mesenteric adipose tissue obtained separately with informed consent, total RNA was extracted using TRIzol (Invitrogen Life Technologies). After treatment with DNase I, total RNAs were purified using mRNA Purification Kits (Amersham Biosciences) according to the manufacturer's instructions. All poly(A)<sup>+</sup> RNAs were pooled from at least two normal donors. T7-based RNA amplification was carried out as described previously.<sup>18</sup> Using 2 µg of poly(A)<sup>+</sup> RNA as starting material, we performed one round of amplification to obtain 100–200 µg of amplified RNA (aRNA) for each tissue.

### 2.2. cDNA microarrays

We constructed a "genome-wide" cDNA microarray consisting of 23,040 cDNAs selected from the UniGene database (build #131) of the National Center for Biotechnology Information (NCBI), as described previously.<sup>7–9</sup> Briefly, the cDNAs were amplified by RT-PCR using poly(A)<sup>+</sup> RNA isolated from various human organs as templates; the lengths of the amplicons ranged from 200 to 1,100 bp excluding repetitive or poly(A) sequences. The PCR products were spotted on type 7 glass slides using a Microarray Spotter, Generation III (Amersham Biosciences); 4608 genes were spotted in duplicate on a single slide. We prepared five different sets of slides (i.e., 23,040 genes total), on each of which the same 52 housekeeping genes and 2 negative-control genes were spotted as well.

### 2.3. Labeling and hybridization

A fluorescent cDNA probe was prepared from each aRNA in the manner described previously;<sup>9</sup> 12.5 µg aliquots of aRNA from individual tissues and a mixture of equivalent aRNA from all 29 tissues as a universal control were labeled respectively with Cy5-dCTP and Cy3-dCTP dyes. Hybridization and washing were performed according to protocols described previously<sup>18</sup> except that all processes were carried out with an Auto Slide Processor (Amersham Biosciences).

### 2.4. Quantification of signals

The intensity of each hybridization signal was calculated photometrically by the ArrayVision computer pro-

gram (Amersham Biosciences) and background intensity was subtracted. Normalization of each Cy3- and Cy5-fluorescence intensity was performed using averaged signals from the 52 housekeeping genes. A cut-off value for each expression level was automatically calculated according to background fluctuation. When both Cy3 and Cy5 signal intensities were lower than the cut-off values, we assessed expression of the corresponding gene in that sample as absent. In cases where only the Cy3 signal was lower than the cut-off value, we replaced its signal with the cut-off value and calculated the relative expression ratio as Cy5/cut-off. If only the Cy5 signal was lower than the cut-off value, we calculated cut-off/Cy3. For other genes, we calculated Cy5/Cy3 as the relative expression ratio.

### 2.5. Cluster analysis of 29 normal tissues

We applied hierarchical clustering analysis to both genes and samples using web-available software ("Cluster and TreeView") written by M. Eisen (<http://genome-www5.stanford.edu/MicroArray/SMD/restech.html>).<sup>19</sup> To obtain reproducible clusters, we excluded genes where both the Cy3 and Cy5 signals were lower than the cut-off value, and selected only genes for which the Cy5/Cy3 ratio was greater than five in at least one sample. Before the clustering algorithm was applied, the fluorescence ratio for each spot was log-transformed and the data for each sample were centered to remove experimental biases. The selection criteria left 3,408 genes for the clustering analysis.

### 2.6. Northern analysis

A 0.5 µg aliquot of poly(A)<sup>+</sup> RNA from each of the 29 normal human tissues was electrophoresed on a 1% agarose gel containing 1 × MOPS buffer and 2% formaldehyde, and was transferred to a nylon membrane. Hybridization with random-primed <sup>32</sup>P-labelled probes was carried out according to the manufacturer's instructions (Clontech) using the same RT-PCR products as those spotted onto the cDNA microarray slides.

## 3. Results

### 3.1. Classification of genes

We analyzed gene-expression profiles in 25 adult and 4 fetal human tissues on our cDNA microarray, using a mixture of poly(A)<sup>+</sup> RNAs from all 29 tissues as a control to calculate the relative expression ratio (Cy5/Cy3) of each of the 23,040 genes on the array (see Materials and Methods). The numbers of genes that showed Cy5/Cy3 ratios of 5 or higher or those of 10 or higher in individual tissues are summarized in Fig. 1. Of the 4,080 genes that were highly expressed in only one or a few tissues, 1,163 were expressed with a ratio of > 10 in only one tissue. Liver, placenta, skeletal muscle, and

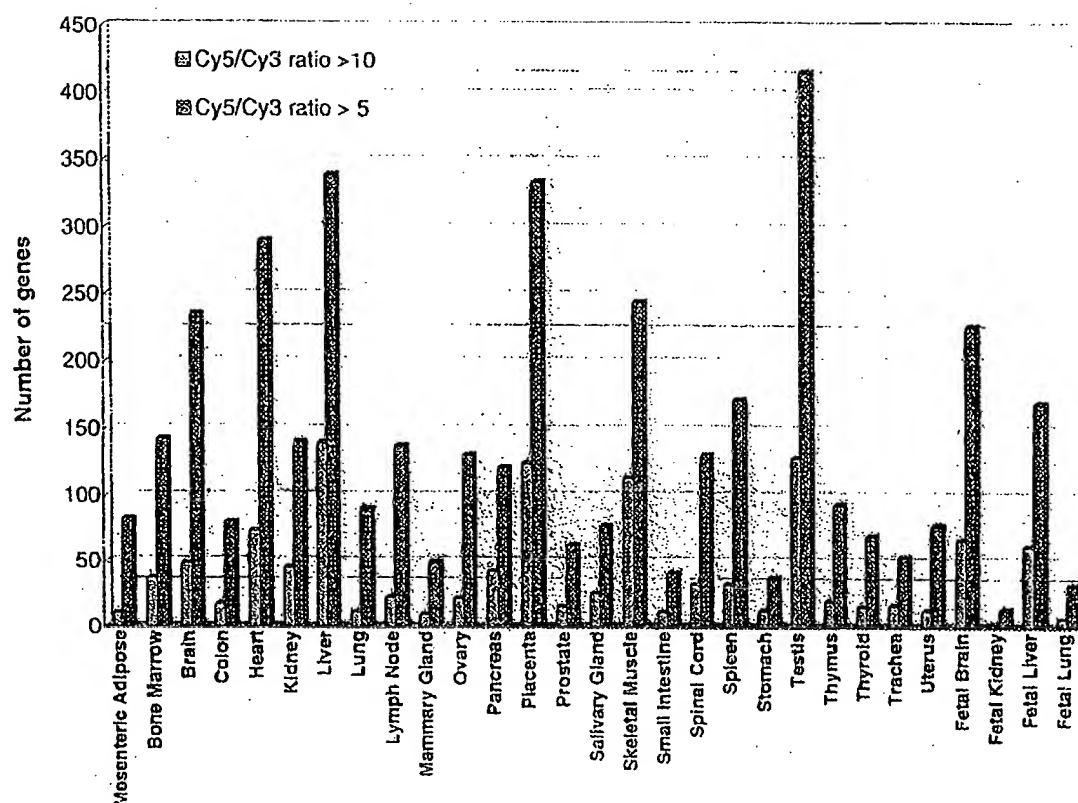


Figure 1. Distribution of genes highly expressed in 29 normal human tissues.

testis each contained more than 100 genes that showed at least a tenfold higher level of expression there than in other tissues.

Table 1 lists the 30 genes showing the biggest differences in expression levels in a given tissue as compared with the mixture of mRNAs from all 29 tissues. The list includes genes that were previously reported to be related to or responsible for a disease. For example, among the 30 genes highly expressed in kidney, SLC3A1 (solute carrier family 3 member 1, with a Cy5/Cy3 ratio of 26.6), is involved in the transport of dibasic and neutral forms of cystine, and its mutations cause cystinuria type 1.<sup>20,21</sup> In heart, mutant forms of TNNT2 (cardiac troponin T), TNNT2 (cardiac troponin T2), ACTC (alpha cardiac actin), and MYH7 (beta cardiac myosin heavy chain) are capable of causing hypertrophic cardiomyopathy.<sup>22-25</sup> Their Cy5/Cy3 ratios were 56.1, 43.3, 31.4, and 18.6, respectively. A comparison with data from the Body Map database, which also contains ESTs from human tissues, indicated that many of the tissue-specific genes listed here were also indicated by Body Map to be expressed highly in mesenteric adipose tissue, adult liver, adult lung, skeletal muscle, fetal brain, fetal liver, and fetal lung; those genes are shown in bold letters in Table 1.

### 3.2. Clustering analysis of gene-expression patterns in normal human tissues

We applied an algorithm for hierarchical clustering of the expression profiles of all 29 tissues, using 3,408 genes whose expression ratios were 5 or higher in at least one tissue (Fig. 2). The colored bars in Fig. 2 indicate clusters of genes that were coordinately expressed in a tissue-specific manner. Fetal brain, adult brain, and spinal cord were clustered into the same terminal branch, indicating that a large number of genes are commonly expressed in those neuronal tissues. In this cluster we found genes encoding neuronal cytoskeleton proteins, neurofilament components (NEFL and NEF3), microtubule components (TUBB, TUBB1, and TUBA3, tubulin polypeptides), MAPT (microtubule associated protein tau), neuroendocrine-related proteins (SGNE1, RTN1, RTN3, CADPS), and proteins involved in neuronal vesicle trafficking (KIF5C, SH3GL2, STXBP1). Also included in this cluster were GPM6B, GPM6A, NR-CAM, ASTN, and CTNND, which encode neural cell adhesion molecules, and neurotransmission-related proteins GRM3 (glutamate receptor) and GAD1 (glutamate decarboxylase 1) (Fig. 3a).

Spleen, lymph node, thymus and bone marrow con-

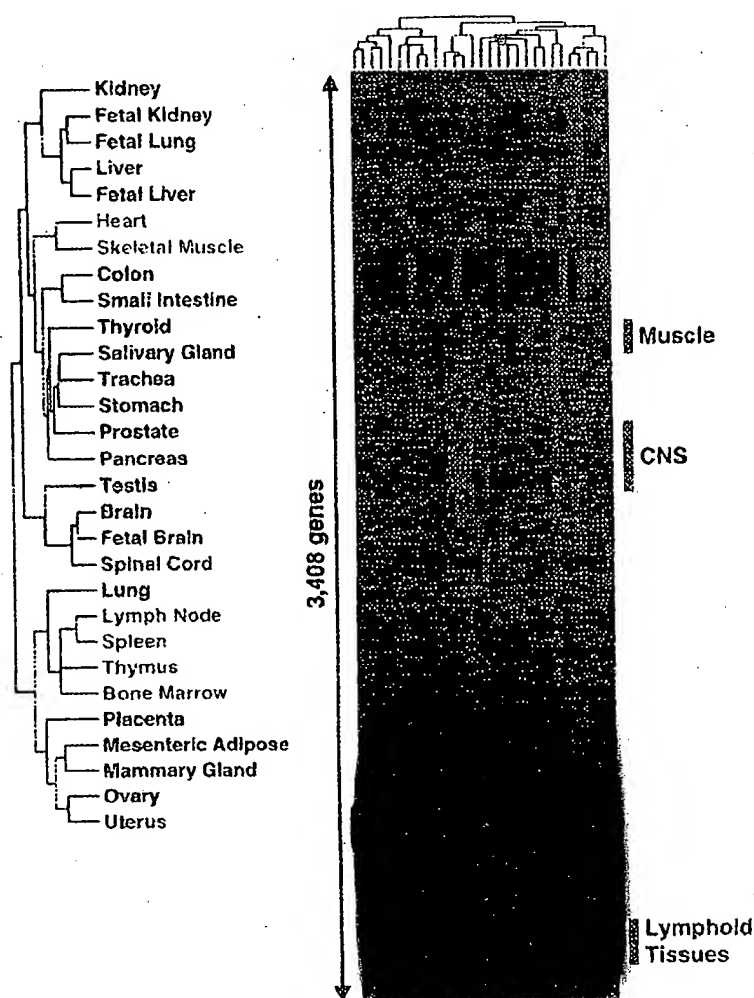


Table 1. Continued.

[illegible]

The table lists the 30 genes in each tissue showing the highest Cy3/Cy3 ratios, except for fetal kidney where we identified only 13 genes with ratios  $> 5$ . Gene identifiers were assigned in our laboratory. In the columns for adipose tissue, colon, adult liver, fetal brain, and skeletal muscle, bold type indicates genes for which our data overlapped the Body Map database. The Hs. numbers were retrieved from the Unigene Database (build #131, NCBI).





**Figure 2.** Hierarchical clustering of gene-expression profiles. Of 23,040 genes on the microarray, 3,408 were selected for clustering analysis by the criteria described in Materials and Methods. The dendrogram at the left lists the tissues studied and provides a measure of the relatedness of gene expression in each sample. In the matrix, each row represents a single cDNA on the microarray, whereas each column corresponds to a separate sample. Red and green colors indicate genes with high and low expression respectively, mRNA compared to the reference mixture of equivalent aRNA from all 29 tissues. Black and gray indicate unchanged expression and no expression, respectively. Colored bars on the right of the matrix indicate clusters of tissue-specific genes of Muscle (heart and skeletal), CNS (adult brain, fetal brain, spinal cord), and Lymphoid tissues (bone marrow, lymph node, spleen, thymus).

stituted another cluster, which included cell-adhesion molecules or receptors (SELL, ITGA4, CD37, CXCR4, IL2RG, CD53, HEM1, and LAIR1), and also PLCB, ARHGDIB, and HCLS1 whose products play important roles in signal transduction. In addition, we found the CORO1A and WASPIP genes, encoding leukocytic cytoskeleton-related proteins, in this cluster (Fig. 3b). Striated-muscle tissues (skeletal muscle and heart) also clustered together. Genes encoding myofibrillar component-proteins (ACTN2, MYH6, MYH3, MYH7, TTN, MYL2, TCAP, MYOM1, and MYOM2) were included in this cluster. ATP2A, CASQ2, and

HRC, encoding calcium-regulating proteins, showed high expression in both types of muscle. Genes encoding enzymes playing important roles in muscles, PFKM, COX6A2, and CKM, were included also (Fig. 3c).

Furthermore, we identified 371 tissue-specific genes that had Cy5/Cy3 ratios  $> 5$  in a single tissue but ratios of  $< 2$  in all other tissues examined (Table 2). In the case of brain, kidney, liver, and lung, we selected tissue-specific genes according the following criteria: Cy5/Cy3 ratios  $> 5$  in either or both of the fetal and adult tissues and ratios  $< 2$  in the remaining 27 tissues. Very few genes were expressed in both fetal and

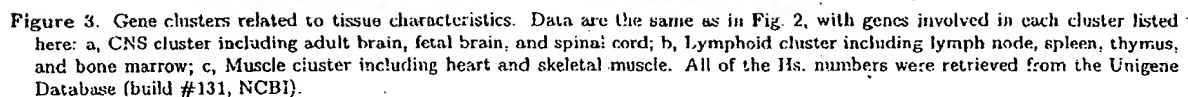


Table 2. Number of tissue-specific genes identified in each tissue.

Tissue	Genes (Cy5/Cy3 > 10)	Genes (Cy5/Cy3 > 5)
Kidney	15 (2)	25 (5)
Adult Kidney	10 (2)	17 (3)
Fetal Kidney	1 (0)	2 (0)
Both	1 (0)	2 (0)
Lung	6 (2)	12 (2)
Adult Lung	0	2 (0)
Fetal Lung	0	0
Both	2 (1)	5 (2)
Liver	55 (7)	100 (29)
Adult Liver	15 (1)	45 (17)
Fetal Liver	3 (2)	6 (4)
Both	12 (0)	33 (4)
Muscle	9 (1)	19 (8)
Heart	50 (17)	90 (40)
Skeletal Muscle	14 (5)	32 (14)
Both		
CNS	89 (13)	119 (53)
Brain	8 (3)	25 (14)
Adult Brain	3 (3)	18 (12)
Fetal Brain	4 (1)	25 (12)
Both	0	4 (3)
Spinal Cord	0	10 (2)
Al		
Lymphoid tissues	0	10 (5)
Lymph Node	1 (1)	6 (4)
Spleen	4 (1)	16 (9)
Thymus	11 (1)	18 (4)
Bone Marrow	0	1 (0)
Al		
Adipose tissues	0	5 (3)
Mesenteric Adipose	2 (3)	5 (0)
Mammary Gland	0	4 (1)
Both		
Others	0	3 (0)
Colon	1 (0)	3 (0)
Small Intestine	9 (4)	14 (7)
Thyroid	5 (1)	8 (1)
Salivary Gland	8 (1)	4 (2)
Trachea	5 (1)	5 (1)
Stomach	8 (1)	14 (9)
Prostate	14 (2)	20 (9)
Pancreas	78 (42)	190 (89)
Testis	47 (10)	79 (19)
Placenta	1 (0)	5 (4)
Ovary	0	1 (1)
Uterus		

Parentheses indicate numbers of qualifying genes of unknown function in the given tissue.

adult tissues.

Many of the tissue clusters were distinguished by specific expression of metabolically related groups of genes. For example, in the muscle cluster, 14 genes showed Cy5/Cy3 ratios > 10 in both skeletal muscle and heart. This set included genes known to play critical roles in muscle, including tropomyosin, proteins of the troponin complex, and myosin heavy polypeptide. In the adipose-tissue cluster, perilipin (PLIN) showed high expression in both mesenteric adipose tissue and mammary gland; testis, placenta, skeletal muscle, liver, and brain tissues each expressed more than 80 genes with Cy5/Cy3 ratios > 5.

### 3.3. Verification of cDNA microarray data

To verify the reliability of our cDNA microarray analysis, we randomly selected ten genes (one of known function and the other nine unknown) which had shown Cy5/Cy3 ratios of more than 10 in a single tissue, and we performed northern blotting using those genes as probes. As Fig. 4 shows, each of these ten genes was confirmed to be expressed in a tissue-specific manner, a result that supported the reliability and rationality of our strategy to identify tissue-specific genes on cDNA microarrays.

## 4. Discussion

We examined a total of 668,160 measurements of gene expression in 25 adult and 4 fetal tissues, by means of a cDNA microarray. This is the first study to apply that technique for analyzing gene expression profiles of normal human tissues on a large scale. By doing this, we established an expression database for sets of genes expressed specifically in a certain tissue(s). Among them we found many genes that were related to various diseases, an indication that genes expressed exclusively in one particular tissue or predominantly expressed in only a few tissues may play crucial roles in tissue-specific function. For example, as shown in Table 3, a defect of the PYGM (muscle glycogen phosphorylase) gene, which showed skeletal muscle-specific expression in our experiments (Cy5/Cy3 ratio = 34.1), causes a typical metabolic myopathy known as glycogen storage disease V (McArdle's disease).<sup>26</sup> Similarly, a defect in AMPD1 (adenosine monophosphate deaminase), which was highly expressed (Cy5/Cy3 ratio = 30.3) only in skeletal muscle, is responsible for myoadenylate deaminase deficiency.<sup>27</sup> In addition, we found that three genes correlated with hypertrophic cardiomyopathy, TNNT3, TNNT2, and ACTC,<sup>22-24</sup> were heart-specific (Cy5/Cy3 = 56.1, 43.3, and 31.4, respectively). Moreover, the MYH7 (beta-myosin heavy-chain) gene, included in the muscle cluster defined by our clustering analyses (Fig. 3c), is responsible for hypertrophic cardiomyopathy (HCM); mutations in the MYH7 gene are associated also with skeletal muscle changes characteristic of central core disease.<sup>25</sup>

SFTPB (pulmonary surfactant protein B), a lung-specific gene according to our results (Cy5/Cy3 ratio = 10.3 and 11.2 in adult and fetal lung, respectively), is responsible for congenital alveolar proteinosis.<sup>28</sup> Similarly, we detected specificity of TPO (thyroid peroxidase) to thyroid tissue (Cy5/Cy3 ratio = 24.1). TPO is a key enzyme in the synthesis of thyroid hormones; defects of the TPO gene can lead to congenital hypothyroidism due to a total iodide organification defect (TIOD).<sup>29</sup> Recently, Blackshaw and coworkers identified genes that were specific to or highly enriched in rods using gene-expression profile by the SAGE technique, and they demonstrated that many of candidate retinal disease genes were identified on the basis of these expression profiles and integrating these data with *in silico* mapping.<sup>30</sup> These results suggest that our study may be useful for cloning genes associated with disease involved in specific organs where gene-specific expression of genes is observed.

It is important to identify tissue-specific genes in the interest of cancer research as well. Using a differential display method, Ozaki et al. demonstrated that a pancreas-specific gene, pancpin, was down-regulated or mutated in primary pancreatic cancers.<sup>11</sup> All these lines of evidence indicate that obtaining sets of genes expressed in specific tissues may reveal genes that have not yet been

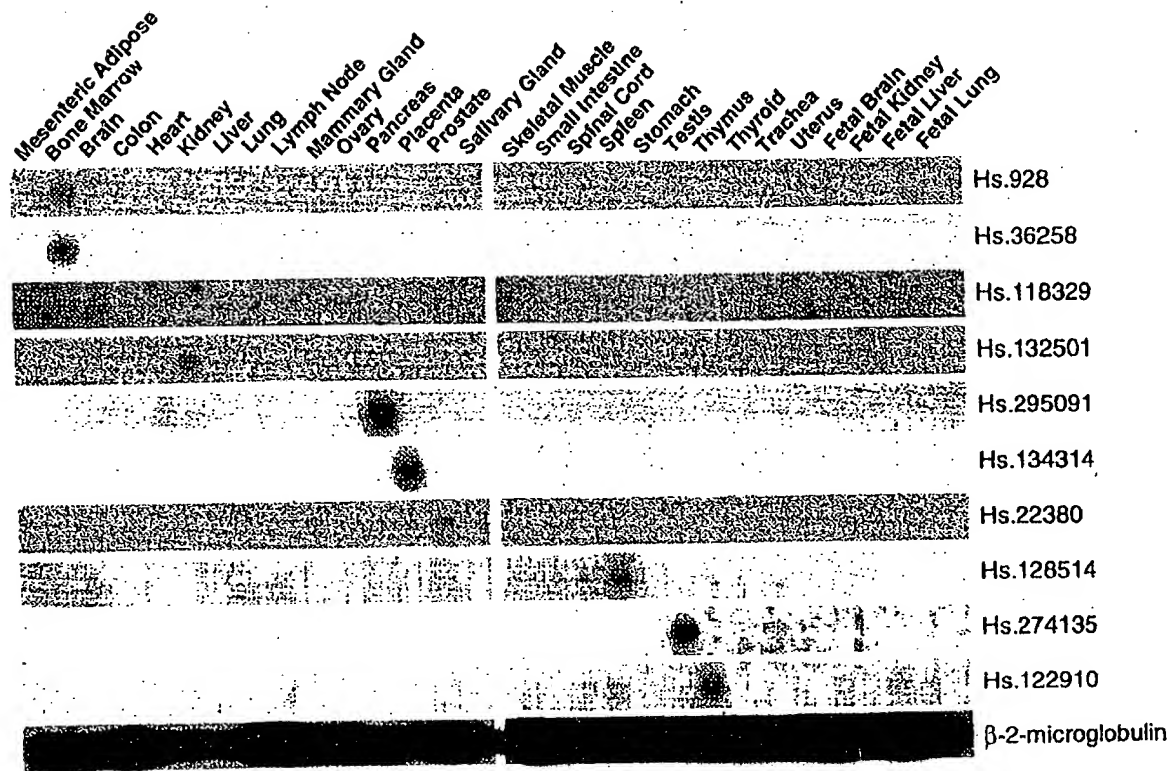


Figure 4. Confirmation of reliability of the microarray data. The tissue-specific expression of ten genes whose expression ratios were increased more than tenfold in a single tissue and less than twofold in the remaining tissues were confirmed by northern analyses. Hs. 928, PRTN3 (Cy5/Cy3 ratio = 25.6 in bone marrow); Hs. 36258, EST (Cy5/Cy3 ratio = 33.0 in bone marrow); Hs. 118329, EST (Cy5/Cy3 ratio = 10.3 in adult brain); Hs. 132501, EST (Cy5/Cy3 ratio = 15.7 in adult kidney); Hs. 295091, EST (Cy5/Cy3 ratio = 42.04 in pancreas); Hs. 134314, EST, (Cy5/Cy3 ratio = 27.8 in placenta); Hs. 22380, EST, (Cy5/Cy3 ratio = 10.28 in prostate); Hs. 128514, EST, (Cy5/Cy3 ratio = 11.9 in spleen); Hs. 274135, DKFZp434K1815 (Cy5/Cy3 ratio = 12.5 in testis); Hs. 122910, EST (Cy5/Cy3 ratio = 11.2 in thymus). All of the Hs. numbers were retrieved from the Unigene Database (build #131, NCBI). Beta-2-microglobulin served as a control.

identified as responsible for human disease.

Transcriptional regulation of genes in a given tissue is controlled by a combination of several mechanisms; e.g. changes in chromatin structure mediated by the association of histone acetylases and deacetylases,<sup>31</sup> binding of transcriptional regulators to an enhancer and/or a suppressor,<sup>32</sup> binding of transcription factors to a promoter region, and methylation of CpG islands in promoter sequences.<sup>33</sup> However, the precise mechanisms of cooperation among these elements are not well understood.<sup>34,35</sup> To address this issue, we identified several potential regulatory elements of liver-specific transcription on the basis of tissue-specific expression (data not shown). Comparison of promoter sequences among a large number of tissue-specific genes would identify their common cis-elements whose core sequence might be useful for identifying novel transcription-related factors in each tissue and for use as promoters in vectors designed for gene therapy.<sup>36</sup>

Using a clustering algorithm, we found that three of the fetal tissues examined (kidney, liver, and lung) were clustered in the same branch. This implies that gene expression patterns among the three tissues might be significantly similar during fetal life. Sell et al. reported that the albumin gene was expressed in both kidney and lung of fetal rat;<sup>37</sup> we confirmed a relatively high level of expression of albumin mRNA in human fetal kidney and fetal lung (data not shown). However, in our analysis, expression profiles of genes in adult and fetal lung tissues fell into different branches an indication that fetal lung, with no air respiration, functions quite differently from adult lung. We conclude that hierarchical clustering of normal tissues on the basis of gene expression can elucidate characteristic features of each tissue.

Finally, our information should have relevance to cancer therapy. We have already carried out expression profiling in many types of cancer including ovarian,<sup>18</sup> hepatocellular,<sup>9</sup> colorectal,<sup>38</sup> and esophageal<sup>39</sup> carcino-

Table 3. List of tissue-specific genes related to human diseases in this study.

Gene ID	Cy5/Cy3	Symbol	Gene Name	Function	Related disease	Ref
<b>Skeletal muscle</b>						
A6921	34.1	PYGM	muscle glycogen phosphorylation	allosteric enzyme in carbohydrate metabolism	Glycogen storage disease V	26
A3455	30.3	AMPD1	adenosine monophosphate deaminase	energy metabolism.	myoadenylate deaminase deficiency	27
<b>Heart</b>						
C4166	56.1	TNNI3	cardiac troponin I	tropomyosin-binding subunit of troponin	Hypertrophic cardiomyopathy	22
A7164	43.3	TNNI2	cardiac troponin T2	tropomyosin-binding subunit of troponin	Hypertrophic cardiomyopathy	23
B2663	31.4	ACTC	alpha cardiac actin	cell motility	Hypertrophic cardiomyopathy	24
A0840	18.6	MYH7	beta-myosin heavy chain 7	muscle contraction	Hypertrophic cardiomyopathy	25
<b>Lung</b>						
A1739	10.8	SFTPB	pulmonary surfactant protein B	promote alveolar stability	congenital alveolarproliferosis	28
<b>Thyroid</b>						
A3240	24.1	TPO	thyroid peroxidase	iodination and coupling of the hormonogenic tyrosines in thyroglobulin	total iode organification defect	29

Gene ID represents the ID in our laboratory. Ref indicates a reference concerning each related disease.

mas using "genome-wide" cDNA microarrays. Those expression profiles provide useful information for identifying genes that are specifically expressed in cancer cells and whose products would be good candidates as molecular targets for novel drugs. Furthermore, if these candidate genes are not expressed in vital tissues, such drugs may be less likely to cause adverse effects. Thus, the gene expression profiles we have now established in normal tissues, especially those of tissue-specific genes, will serve as references for estimating the risk of adverse effects and should help to guide the selection of molecular targets for the development of novel therapeutic drugs.

**Acknowledgements:** We are grateful to Drs. Tetsushi Yada and Natsuhiko Ichinose of Human Genome Center, Institute of Medical Science, University of Tokyo for promoter analysis *in silico*. We also thank Hiroko Bando, Noriko Nemoto, Noriko Sudo, Kie Naito, and Yukiko Tsuno for fabrication of the cDNA microarray, and Dr. Suguru Hasegawa for helpful discussion. This work was supported in part by Research for the Future Program Grant #00L01402 from the Japan Society for the Promotion of Science.

## References

- Lander, E. S., Linton, L. M., Birren, B. et al. 2001, Initial sequencing and analysis of the human genome, *Nature*, 409, 860-921.
- Venter, J. C., Adams, M. D., Myers, E. W. et al. 2001, The sequence of the human genome, *Science*, 291, 1304-1351.
- Adams, M. D., Kelley, J. M., Cocayne, J. D. et al. 1991, Complementary DNA sequencing: expressed sequence tags and human genome project, *Science*, 252, 1651-1656.
- DeRisi, J. L., Iyer, V. R., and Brown, P. O. 1997, Exploring the metabolic and genetic control of gene expression on a genomic scale, *Science*, 278, 680-686.
- Behr, M. A., Wilson, M. A., Gill, W. P. et al. 1999, Comparative genomics of BCG vaccines by whole-genome DNA microarray, *Science*, 284, 1520-1523.
- Scherf, U., Ross, D. T., Waltham, M. et al. 2000, A gene expression database for the molecular pharmacology of cancer, *Nat. Genet.*, 24, 236-244.
- Golub, T. R., Slonim, D. K., Tamayo, P. et al. 1999, Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, *Science*, 286, 531-537.
- Alizadeh, A. A., Eisen, M. B., Davis, R. E. et al. 2000, Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling, *Nature*, 403, 503-511.
- Okabe, H., Satoh, S., Kato, T. et al. 2001, Genome-wide analysis of gene expression in human hepatocellular carcinomas using cDNA microarray: identification of genes involved in viral carcinogenesis and tumor progression, *Cancer Res.*, 61, 2129-2137.
- Ozaki, K., Kuroki, T., Hayashi, S., and Nakamura, Y. 1996, Isolation of three testis-specific genes (TSA303, TSA806, TSA903) by a differential mRNA display method, *Genomics*, 36, 316-319.
- Ozaki, K., Nagata, M., Suzuki, M. et al. 1998, Isolation

- and characterization of a novel human pancreas-specific gene, *pancpin*, that is down-regulated in pancreatic cancer cells, *Genes Chromosomes Cancer*, **22**, 179-185.
12. Ozaki, K., Nagata, M., Suzuki, M. et al. 1998, Isolation and characterization of a novel human lung-specific gene homologous to lysosomal membrane glycoproteins 1 and 2: significantly increased expression in cancers of various tissues, *Cancer Res.*, **58**, 3499-3503.
  13. Kawamoto, S., Matsumoto, Y., Mizuno, K., Okubo, K., and Matsubara, K. 1996, Expression profiles of active genes in human and mouse livers, *Gene*, **174**, 151-158.
  14. Maeda, K., Okubo, K., Shimomura, I., Mizuno, K., Matsuzawa, Y., and Matsubara, K. 1997, Analysis of an expression profile of genes in the human adipose tissue, *Gene*, **190**, 227-235.
  15. Kawamoto, S., Yoshii, J., Mizuno, K. et al. 2000, BodyMap: a collection of 3' ESTs for analysis of human gene expression information, *Genome Res.*, **10**, 1817-1827.
  16. Soejima, H., Kawamoto, S., Akai, J. et al. 2001, Isolation of novel heart-specific genes using the BodyMap database, *Genomics*, **74**, 115-120.
  17. Moreno, J. C., Pauws, E., van Kampen, A. H., Jedlickova, M., de Vijlder, J. J., and Ris-Stalpers, C. 2001, Cloning of tissue-specific genes using serial analysis of gene expression and a novel computational subtraction approach, *Genomics*, **75**, 70-76.
  18. Ono, K., Tanaka, T., Tsunoda, T. et al. 2000, Identification by cDNA microarray of genes involved in ovarian carcinogenesis, *Cancer Res.*, **60**, 5007-5011.
  19. Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. 1998, Cluster analysis and display of genome-wide expression patterns, *Proc. Natl. Acad. Sci. USA*, **95**, 14863-14868.
  20. Calonge, M. J., Gasparini, P., Chillaron, J. et al. 1994, Cystinuria caused by mutations in rBAT, a gene involved in the transport of cystine, *Nat. Genet.*, **6**, 420-425.
  21. Pras, E., Arber, N., Aksentijevich, I. et al. 1994, Localization of a gene causing cystinuria to chromosome 2p, *Nat. Genet.*, **6**, 415-419.
  22. Kimura, A., Harada, H., Park, J. E. et al. 1997, Mutations in the cardiac troponin I gene associated with hypertrophic cardiomyopathy, *Nat. Genet.*, **16**, 379-382.
  23. Thierfelder, L., Watkins, H., MacRae, C. et al. 1994, Alpha-tropomyosin and cardiac troponin T mutations cause familial hypertrophic cardiomyopathy: a disease of the sarcomere, *Cell*, **77**, 701-712.
  24. Mogensen, J., Klausen, I. C., Pedersen, A. K. et al. 1999, Alpha-cardiac actin is a novel disease gene in familial hypertrophic cardiomyopathy, *J. Clin. Invest.*, **103**, R39-43.
  25. Fananapazir, L., Dalakas, M. C., Cyran, F., Cohn, G., and Epstein, N. D. 1993, Missense mutations in the beta-myosin heavy-chain gene cause central core disease in hypertrophic cardiomyopathy, *Proc. Natl. Acad. Sci. USA*, **90**, 3993-3997.
  26. Gautron, S., Daegelen, D., Memecier, F., Dubocq, D., Kahn, A., and Dreyfus, J. C. 1987, Molecular mechanisms of McArdle's disease (muscle glycogen phosphorylase deficiency). RNA and DNA analysis, *J. Clin. Invest.*, **79**, 275-281.
  27. Morisaki, T., Gross, M., Morisaki, H., Pongratz, D., Zollner, N., and Holmes, E. W. 1992, Molecular basis of AMP deaminase deficiency in skeletal muscle, *Proc. Natl. Acad. Sci. USA*, **89**, 6457-6461.
  28. Nogee, L. M., de Mello, D. E., Dehner, L. P., and Collen, H. R. 1993, Brief report: deficiency of pulmonary surfactant protein B in congenital alveolar proteinosis, *N. Engl. J. Med.*, **328**, 406-410.
  29. Bikker, H., Vulsma, T., Baas, F., and de Vijlder, J. J. 1995, Identification of five novel inactivating mutations in the human thyroid peroxidase gene by denaturing gradient gel electrophoresis, *Hum. Mutat.*, **6**, 9-16.
  30. Blackshaw, S., Fraioli, R. E., Furukawa, T., and Cepko, C. L. 2001, Comprehensive analysis of photoreceptor gene expression and the identification of candidate retinal disease genes, *Cell*, **107**, 579-589.
  31. Wade, P. A. 2001, Transcriptional control at regulatory checkpoints by histone deacetylases: molecular connections between cancer and chromatin, *Hum. Mol. Genet.*, **10**, 693-698.
  32. Luscher, B. and Larsson, L. G. 1999, The basic region/helix-loop-helix/leucine zipper domain of Myc proto-oncoproteins: function and regulation, *Oncogene*, **18**, 2955-2966.
  33. Ballestar, E. and Wolffe, A. P. 2001, Methyl-CpG-binding proteins. Targeting specific gene repression, *Eur. J. Biochem.*, **268**, 1-6.
  34. Cha, J. Y., Kim, H., Kim, K. S., Hur, M. W., and Ahn, Y. 2000, Identification of transacting factors responsible for the tissue-specific expression of human glucose transporter type 2 isoform gene. Cooperative role of hepatocyte nuclear factors 1alpha and 3beta, *J. Biol. Chem.*, **275**, 18358-18365.
  35. Bhavsar, P. K., Dellow, K. A., Yacoub, M. H., Brand, N. J., and Barton, P. J. 2000, Identification of cis-acting DNA elements required for expression of the human cardiac troponin I gene promoter, *J. Mol. Cell. Cardiol.*, **32**, 95-108.
  36. Nettelbeck, D. M., Jerome, V., and Muller, R. 2000, Gene therapy: designer promoters for tumour targeting, *Trends Genet.*, **16**, 174-181.
  37. Sell, S., Longley, M. A., and Boulter, J. 1985, alpha-Fetoprotein and albumin gene expression in brain and other tissues of fetal and adult rats, *Brain Res.*, **354**, 49-53.
  38. Kitabara, O., Furukawa, Y., Tanaka, T. et al. 2001, Alterations of gene expression during colorectal carcinogenesis revealed by cDNA microarrays after laser-capture microdissection of tumor tissues and normal epithelia, *Cancer Res.*, **61**, 3544-3549.
  39. Kihara, C., Tsunoda, T., Tanaka, T. et al. 2001, Prediction of sensitivity of esophageal tumors to adjuvant chemotherapy by cDNA microarray analysis of gene-expression profiles, *Cancer Res.*, **61**, 6474-6479.

## Correlation between Protein and mRNA Abundance in Yeast

STEVEN P. GYGI, YVAN ROCHON, B. ROBERT FRANZA, AND RUEDI AEBERSOLD\*

Department of Molecular Biotechnology, University of Washington, Seattle, Washington 98195-7730

Received 5 October 1998/Returned for modification 11 November 1998/Accepted 2 December 1998

We have determined the relationship between mRNA and protein expression levels for selected genes expressed in the yeast *Saccharomyces cerevisiae* growing at mid-log phase. The proteins contained in total yeast cell lysate were separated by high-resolution two-dimensional (2D) gel electrophoresis. Over 150 protein spots were excised and identified by capillary liquid chromatography-tandem mass spectrometry (LC-MS/MS). Protein spots were quantified by metabolic labeling and scintillation counting. Corresponding mRNA levels were calculated from serial analysis of gene expression (SAGE) frequency tables (V. E. Velculescu, L. Zhang, W. Zhou, J. Vogelstein, M. A. Basrai, D. E. Bassett, Jr., P. Hieter, B. Vogelstein, and K. W. Kinzler, *Cell* 88:243-251, 1997). We found that the correlation between mRNA and protein levels was insufficient to predict protein expression levels from quantitative mRNA data. Indeed, for some genes, while the mRNA levels were of the same value the protein levels varied by more than 20-fold. Conversely, invariant steady-state levels of certain proteins were observed with respective mRNA transcript levels that varied by as much as 30-fold. Another interesting observation is that codon bias is not a predictor of either protein or mRNA levels. Our results clearly delineate the technical boundaries of current approaches for quantitative analysis of protein expression and reveal that simple deduction from mRNA transcript analysis is insufficient.

The description of the state of a biological system by the quantitative measurement of the system constituents is an essential but largely unexplored area of biology. With recent technical advances including the development of differential display-PCR (21), of cDNA microarray and DNA chip technology (20, 27), and of serial analysis of gene expression (SAGE) (34, 35), it is now feasible to establish global and quantitative mRNA expression profiles of cells and tissues in species for which the sequence of all the genes is known. However, there is emerging evidence which suggests that mRNA expression patterns are necessary but are by themselves insufficient for the quantitative description of biological systems. This evidence includes discoveries of posttranscriptional mechanisms controlling the protein translation rate (15), the half-lives of specific proteins or mRNAs (33), and the intracellular location and molecular association of the protein products of expressed genes (32).

Proteome analysis, defined as the analysis of the protein complement expressed by a genome (26), has been suggested as an approach to the quantitative description of the state of a biological system by the quantitative analysis of protein expression profiles (36). Proteome analysis is conceptually attractive because of its potential to determine properties of biological systems that are not apparent by DNA or mRNA sequence analysis alone. Such properties include the quantity of protein expression, the subcellular location, the state of modification, and the association with ligands, as well as the rate of change with time of such properties. In contrast to the genomes of a number of microorganisms (for a review, see reference 11) and the transcriptome of *Saccharomyces cerevisiae* (35), which have been entirely determined, no proteome map has been completed to date.

The most common implementation of proteome analysis is the combination of two-dimensional gel electrophoresis (2DE)

(isoelectric focusing-sodium dodecyl sulfate [SDS]-polyacrylamide gel electrophoresis) for the separation and quantitation of proteins with analytical methods for their identification. 2DE permits the separation, visualization, and quantitation of thousands of proteins reproducibly on a single gel (18, 24). By itself, 2DE is strictly a descriptive technique. The combination of 2DE with protein analytical techniques has added the possibility of establishing the identities of separated proteins (1, 2) and thus, in combination with quantitative mRNA analysis, of correlating quantitative protein and mRNA expression measurements of selected genes.

The recent introduction of mass spectrometric protein analysis techniques has dramatically enhanced the throughput and sensitivity of protein identification to a level which now permits the large-scale analysis of proteins separated by 2DE. The techniques have reached a level of sensitivity that permits the identification of essentially any protein that is detectable in the gels by conventional protein staining (9, 29). Current protein analytical technology is based on the mass spectrometric generation of peptide fragment patterns that are idiosyncratic for the sequence of a protein. Protein identity is established by correlating such fragment patterns with sequence databases (10, 22, 37). Sophisticated computer software (8) has automated the entire process such that proteins are routinely identified with no human interpretation of peptide fragment patterns.

In this study, we have analyzed the mRNA and protein levels of a group of genes expressed in exponentially growing cells of the yeast *S. cerevisiae*. Protein expression levels were quantified by metabolic labeling of the yeast proteins to a steady state, followed by 2DE and liquid scintillation counting of the selected, separated protein species. Separated proteins were identified by in-gel tryptic digestion of spots with subsequent analysis by microspray liquid chromatography-tandem mass spectrometry (LC-MS/MS) and sequence database searching. The corresponding mRNA transcript levels were calculated from SAGE frequency tables (35).

This study, for the first time, explores a quantitative comparison of mRNA transcript and protein expression levels for a relatively large number of genes expressed in the same metabolic state. The resultant correlation is insufficient for predic-

\* Corresponding author. Mailing address: Department of Molecular Biotechnology, Box 357730, University of Washington, Seattle, WA 98195-7730. Phone: (206) 221-4196. Fax: (206) 685-7301. E-mail: medi@u.washington.edu.



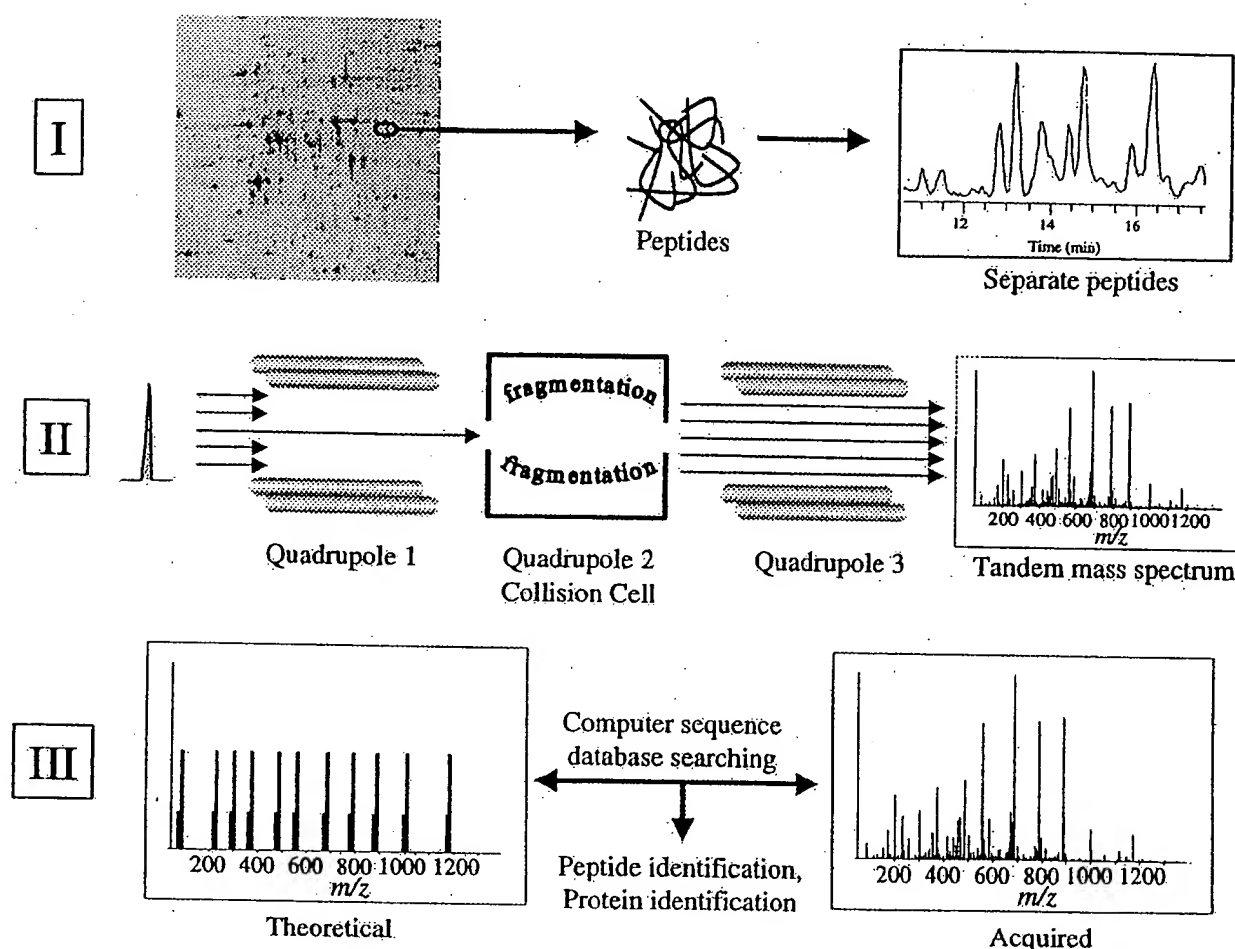


FIG. 1. Schematic illustration of proteome analysis by 2DE and mass spectrometry. In part I, proteins are separated by 2DE, stained spots are excised and subjected to in-gel digestion with trypsin, and the resulting peptides are separated by on-line capillary high-performance liquid chromatography. In part II, a peptide is shown eluting from the column in part I. The peptide is ionized by electrospray ionization and enters the mass spectrometer. The mass of the ionized peptide is detected, and the first quadrupole mass filter allows only the specific mass-to-charge ratio of the selected peptide ion to pass into the collision cell. In the collision cell, the energized, ionized peptides collide with neutral argon gas molecules. Fragmentation of the peptide is essentially random but occurs mainly at the peptide bonds, resulting in smaller peptides of differing lengths (masses). These peptide fragments are detected as a tandem mass (MS/MS) spectrum in the third quadrupole mass filter where two ion series are recorded simultaneously, one each from sequencing inward from the N and C termini of the peptide, respectively. In part III, the MS/MS spectrum from the selected, ionized peptide is compared to predicted tandem mass spectra computer generated from a sequence database. Provided that the peptide sequence exists in the database, the peptide and, by association, the protein from which the peptide was derived can be identified. Unambiguous protein identification is attained in a single analysis because multiple peptides are identified as being derived from the same protein.

tion of protein levels from mRNA transcript levels. We have also compared the relative amounts of protein and mRNA with the respective codon bias values for the corresponding genes. This comparison indicates that codon bias by itself is insufficient to accurately predict either the mRNA or the protein expression levels of a gene. In addition, the results demonstrate that only highly expressed proteins are detectable by 2DE separation of total cell lysates and that therefore the construction of complete proteome maps with current technology will be very challenging, irrespective of the type of organism.

#### MATERIALS AND METHODS

**Yeast strain and growth conditions.** The source of protein and message transcripts for all experiments was YPH499 (*MATa ura3-52 lys2-801 ade2-101 leu2-Δ1 his3-Δ200 trp1-Δ63*) (30). Logarithmically growing cells were obtained by growing yeast cells to early log phase ( $3 \times 10^6$  cells/ml) in YPD rich medium (YPD supplemented with 6 mM uracil, 4.8 mM adenine, and 24 mM tryptophan) at 30°C (35). Metabolic labeling of protein was accomplished in YPD medium

exactly as described elsewhere (4) with the exception that 1 ml of cells was labeled with 3 mCi to offset methionine present in YPD medium. Protein was harvested as described by Garrels and coworkers (12). Harvested protein was lyophilized, resuspended in isoelectric focusing gel rehydration solution, and stored at -80°C.

**2DE.** Soluble proteins were run in the first dimension by using a commercial slabbed electrophoresis system (Multiphor II; Pharmacia Biotech). Immobilized polyacrylamide gel (IPG) dry strips with nonlinear pH 3.0 to 10.0 gradients (Amersham-Pharmacia Biotech) were used for the first-dimension separation. Forty micrograms of protein from whole-cell lysates was mixed with IPG strip rehydration buffer (8 M urea, 2% Nonidet P-40, 10 mM dithiothreitol), and 250 to 380  $\mu$ l of solution was added to individual lanes of an IPG strip rehydration tray (Amersham-Pharmacia Biotech). The strips were allowed to rehydrate at room temperature for 1 h. The samples were run at 300 V-10 mA-5 W for 2 h, then ramped to 3,500 V-10 mA-5 W over a period of 3 h, and then kept at 3,500 V-10 mA-5 W for 15 to 19 h. At the end of the first-dimension run (60 to 70 kV-h), the IPG strips were reequilibrated for 8 min in 2% (wt/vol) diethylenetriamine in 2% (wt/vol) SDS-6 M urea-30% (wt/vol) glycerol-0.05 M Tris HCl (pH 6.8) and for 4 min in 2.5% iodoacetamide in 2% (wt/vol) SDS-6 M urea-30% (wt/vol) glycerol-0.05 M Tris HCl (pH 6.8). Following reequilibration, the strips were transferred and apposed to 10% polyacrylamide second-dimension gels. Polyacrylamide gels were poured in a casting stand with 10% acrylamide-2.67% piperazine diacrylamide-0.375 M Tris base-HCl (pH 8.8)-0.1% (wt/vol) SDS-0.05%



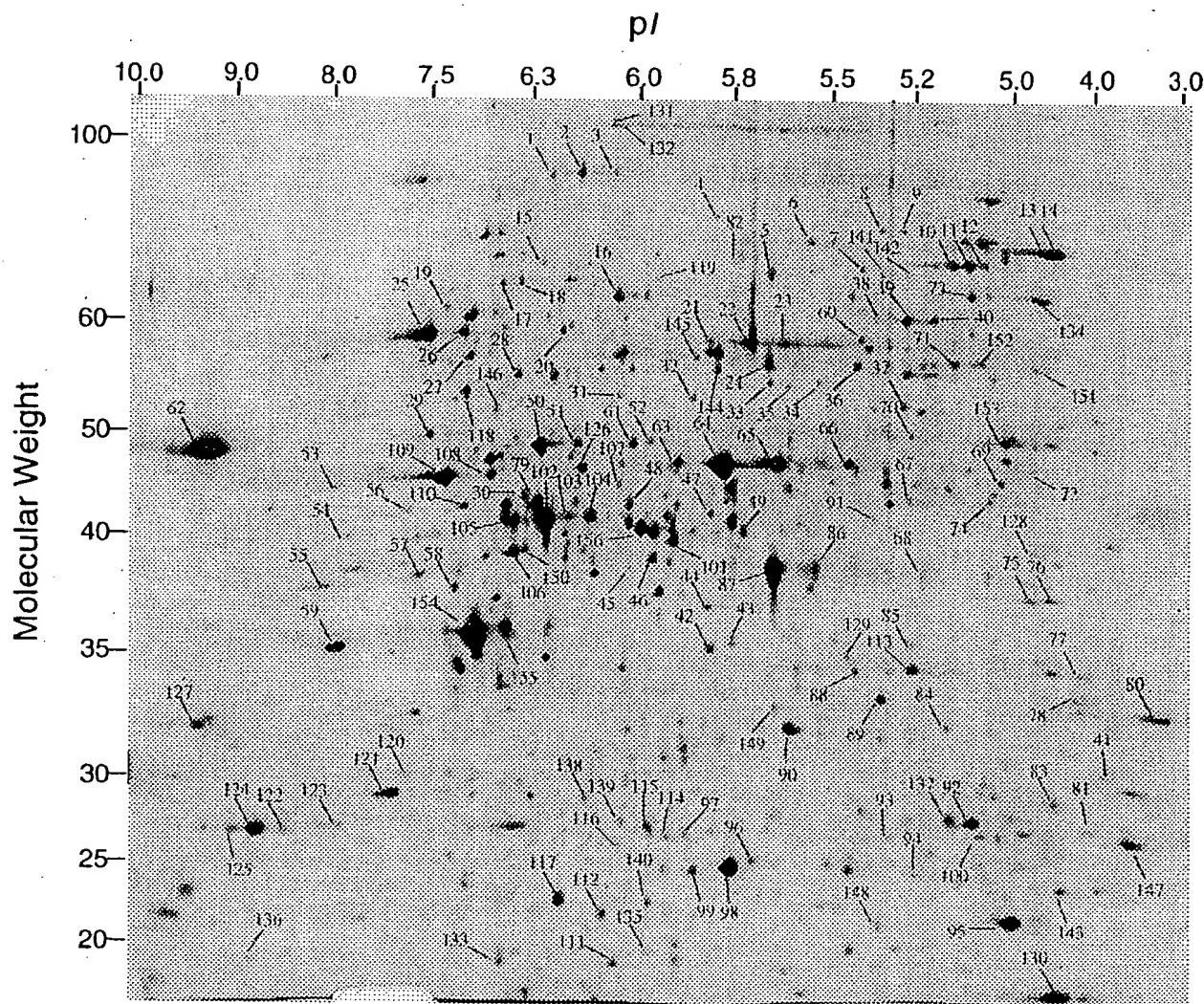


FIG. 2. 2D silver-stained gel of the proteins in yeast total cell lysate. Proteins were separated in the first dimension (horizontal) by isoelectric focusing and then in the second dimension (vertical) by molecular weight sieving. Protein spots (156) were chosen to include the entire range of molecular weights, isoelectric focusing points, and staining intensities. Spots were excised, and the corresponding protein was identified by mass spectrometry and database searching. The spots are labeled on the gel and correspond to the data presented in Table 1. Molecular weights are given in thousands.

(wt/vol) ammonium persulfate–0.05% TEMED (*N,N,N',N'*-tetramethylethylenediamine) in Milli-Q water. The apparatus used to run second-dimension gels was a noncommercial apparatus from Oxford Glycosciences, Inc. Once the IPG strips were apposed to the second-dimension gels, they were immediately run at 50 mA (constant)–500 V–85 W for 20 min, followed by 200 mA (constant)–500 V–85 W until the buffer front line was 10 to 15 mm from the bottom of the gel. Gels were removed and silver stained according to the procedure of Shevchenko et al. (29).

**Protein identification.** Gels were exposed to X-ray film overnight, and then the silver staining and film were used to excise 156 spots of varying intensities, molecular weights, and isoelectric focusing points. In order to increase the detection limit by mass spectrometry, spots were cut out and pooled from up to four identical cold, silver-stained gels. In-gel tryptic digests of pooled spots were performed as described previously (29). Tryptic peptides were analyzed by microcapillary LC-MS with automated switching to MS/MS mode for peptide fragmentation. Spectra were searched against the composite OWL protein sequence database (version 30.2; 250,514 protein sequences) (24a) by using the computer program Sequest (8), which matches theoretical and acquired tandem mass spectra. A protein match was determined by comparing the number of peptides identified and their respective cross-correlation scores. All protein identifications were verified by comparison with theoretical molecular weights and isoelectric points.

**mRNA quantitation.** Velculescu and coworkers have previously generated frequency tables for yeast mRNA transcripts from the same strain grown under the same stated conditions as described herein (35). The SAGE technology is based on two main principles. First, a short sequence tag (15 bp) that contains sufficient information uniquely to identify a transcript is generated. A single tag is usually generated from each mRNA transcript in the cell which corresponds to 15 bp at the 3'-most cutting site for *Nla*III. Second, many transcript tags can be concatenated into a single molecule and then sequenced, revealing the identity of multiple tags simultaneously. Over 20,000 transcripts were sequenced from yeast strain YPH499 growing at mid-log phase on glucose. Assuming the previously derived estimate of 15,000 mRNA molecules per cell (16), this would represent a 1.3-fold coverage even for mRNA molecules present at a single copy per cell and would provide a 72% probability of detecting such transcripts. Computer software which took for input the gene detected, examined the nucleotide sequence, and performed the calculation as described by Velculescu and coworkers (35) was written. In practice, we found that for 21 of 128 (16%) genes examined viable mRNA levels from SAGE data could not be calculated. This was because (i) no CATG site was found in the open reading frame (ORF), (ii) a CATG site was found but the corresponding 10-bp putative SAGE tag was not found in the frequency tables, or (iii) identical putative SAGE tags were present for multiple genes (e.g., *TDH2\_YEAST* and *TDH3\_YEAST*).

TABLE 1. Expressed genes identified from 2D gel in Fig. 2

Mol wt	pI	Spot no.	YPD gene name <sup>a</sup>	Protein abundance (10 <sup>3</sup> copies/cell)	mRNA abundance (copies/cell)	Codon bias
17,259	6.75	133	CPR1	15.2	61.7	0.769
18,702	4.80	83	EGD2	20.1	5.2	0.724
18,726	4.44	147	YKL056C	61.2	88.4	0.831
18,978	5.95	135	YER067W	3.7	6.7	0.118
19,108	5.04	130	YLR109W	94.4	9.7	0.680
19,681	9.08	136	ATP7	11.0	NA <sup>b,c</sup>	0.246
20,505	6.07	111	GUK1	16.5	3.7	0.422
21,444	5.25	148	SAR1	5.4	10.4	0.455
21,583	4.98	95	TSA1	110.6	40.1	0.845
22,602	4.30	80	EFB1	66.1	23.8	0.875
23,079	6.29	112	SOD2	12.6	2.2	0.351
23,743	5.44	137	HSP26	NA <sup>d</sup>	0.7	0.434
24,033	5.97	96	ADK1	17.4	16.4	0.656
24,058	4.43	143	YKL117W	29.2	10.4	0.339
24,353	6.30	140	TFS1	8.1	0.7	0.146
24,662	5.85	99	URA5	25.4	6.0	0.359
24,808	6.33	97	GSP1	26.3	5.2	0.735
24,908	8.73	122	RPS5	18.6	NA <sup>c</sup>	0.899
25,081	4.65	81	MRP8	9.3	NA <sup>c</sup>	0.241
25,960	6.06	116	RPE1	5.8	0.7	0.372
26,378	9.55	127	RPS3	96.8	NA <sup>c</sup>	0.863
26,467	5.18	100	VMA4	10.5	3.7	0.427
26,661	5.84	98	TFI1	NA <sup>d</sup>	NA <sup>c</sup>	0.900
27,156	5.56	93	PRE8	6.9	0.7	0.129
27,334	6.13	115	YHR049W	18.4	2.2	0.520
27,472	5.33	92	YNL010W	31.6	3.7	0.421
27,480	8.95	123	GPM1	10.0	169.4	0.902
27,480	8.95	124	GPM1	231.4	169.4	0.902
27,480	8.95	125	GPM1	7.5	169.4	0.902
27,809	5.97	139	HOR2	5.7	0.7	0.381
27,874	4.46	78	YST1	13.6	52.8	0.805
28,595	4.51	41	PUP2	4.4	0.7	0.147
29,156	6.59	114	YMR226C	14.5	2.2	0.283
29,244	8.40	120	DPM1	5.0	11.2	0.362
29,443	5.91	48	PRE4	3.4	3.7	0.162
30,012	6.39	138	PRB1	21.2	1.5	0.449
30,073	4.63	77	BMH1	14.7	28.2	0.454
30,296	7.94	121	OMP2	67.4	41.6	0.499
30,435	6.34	89	GPP1	70.2	11.2	0.703
31,332	5.57	88	ILV6	13.9	3.0	0.402
32,159	5.46	113	IPP1	63.1	3.7	0.752
32,263	6.00	149	HIS1	22.4	4.5	0.232
33,311	5.35	84	SPE3	15.1	6.7	0.468
34,465	5.60	129	ADE1	8.7	5.2	0.305
34,762	5.32	85	SEC14	10.9	6.0	0.373
34,797	5.85	42	URA1	49.5	8.9	0.237
34,799	6.04	90	BEL1	103.2	81.0	0.875
35,556	5.97	43	YDL124W	6.4	4.5	0.206
35,619	8.41	59	TDH1	69.8	32.7 <sup>e</sup>	0.940
35,650	5.49	68	CAR1	5.2	3.0	0.339
35,712	6.72	117	TDH2	49.6	473.0 <sup>e</sup>	0.982
35,712	6.72	154	TDH2	863.5	473.0 <sup>e</sup>	0.982
35,712	6.72	155	TDH2	79.4	473.0 <sup>e</sup>	0.982
36,272	4.85	128	APA1	8.7	0.7	0.425
36,358	5.05	75	YJR105W	17.6	17.1	0.522
36,358	5.05	76	YJR105W	27.5	17.1	0.522
36,596	6.37	79	ADH2	58.9	260.0 <sup>e</sup>	0.711
36,714	6.30	102	ADH1	746.1	260.0	0.913
36,714	6.30	103	ADH1	17.6	260.0	0.913
36,714	6.30	104	ADH1	61.4	260.0	0.913
36,714	6.30	105	ADH1	52.7	260.0	0.913
37,033	6.23	44	TAL1	44.8	3.7	0.701
37,796	7.36	57	IDH2	29.4	6.7	0.330
37,886	6.49	106	ILV5	76.0	4.5	0.892
38,700	7.83	55	BAT1	30.9	11.2	0.469
38,702	6.24	46	OCR2	NA <sup>d</sup>	2.2	0.326

Continued

TABLE 1—Continued

Mol wt	pI	Spot no.	YPD gene name <sup>a</sup>	Protein abundance (10 <sup>3</sup> copies/cell)	mRNA abundance (copies/cell)	Codon bias
39,477	5.58	86	FBA1	17.8	183.6	0.935
39,477	5.58	87	FBA1	427.2	183.6	0.935
39,540	6.50	150	HOM2	60.3	4.5	0.592
39,561	6.12	156	PSA1	96.4	27.5	0.718
41,158	6.01	49	YNL134C	14.9	1.5	0.316
41,623	7.18	58	BAT2	19.0	8.9	0.250
41,728	7.29	110	ERG10	24.1	4.5	0.543
41,900	5.42	74	TOM40	22.3	2.2	0.375
42,402	6.29	45	CYS3	6.7	8.9	0.621
42,883	5.63	67	DYS1	15.8	5.2	0.526
43,409	6.31	107	SER1	10.5	1.5	0.292
43,421	5.59	91	ERG6	2.2	14.1	0.408
44,174	7.32	56	YBR025C	13.1	6.0	0.684
44,682	4.99	72	TTF1	2.9	39.4	0.834
44,707	7.77	108	PGK1	23.7	165.7	0.897
44,707	7.77	109	PGK1	315.2	165.7	0.897
46,080	6.72	30	CAR2	15.4	NA <sup>c</sup>	0.495
46,383	8.52	53	IDP1	7.7	0.7	0.436
46,553	5.98	47	IDP2	32.4	NA <sup>c</sup>	0.197
46,679	6.39	50	ENO1	35.4	0.7	0.930
46,679	6.39	51	ENO1	6.6	0.7	0.930
46,679	6.39	52	ENO1	2.2	0.7	0.930
46,773	5.82	63	ENO2	15.5	289.1	0.960
46,773	5.82	64	ENO2	635.5	289.1	0.960
46,773	5.82	65	ENO2	93.0	289.1	0.960
46,773	5.82	66	ENO2	31.0	289.1	0.960
47,402	6.09	126	COR1	2.5	0.7	0.422
47,666	8.98	54	AAT2	11.7	6.0	0.338
48,364	5.25	73	WTM1	74.5	13.4	0.365
48,530	6.20	61	MET17	38.1	29.0	0.576
48,904	5.18	69	LYS9	16.2	3.7	0.463
48,987	4.90	153	SUP45	29.6	11.9	0.377
49,727	5.47	70	PRO2	13.6	5.2	0.297
49,912	9.27	62	TEF2	558.5	282.0	0.932
50,444	5.67	35	YDR190C	4.8	2.2	0.228
50,837	6.11	32	YEL047C	3.8	1.5	0.387
50,891	4.59	151	TUB2	11.2	7.4	0.404
51,547	6.80	27	LPD1	18.9	2.2	0.351
52,216	7.25	29	SHM2	19.7	7.4	0.722
52,859	5.54	37	YFR044C	30.2	6.7	0.442
53,798	5.19	71	HXK2	26.5	7.4	0.756
53,803	6.05	145	GYP6	4.4	0.7	0.147
54,403	5.29	39	ALD6	37.7	2.2	0.664
54,403	5.29	40	ALD6	6.6	2.2	0.664
54,502	6.20	31	ADE13	6.3	1.5	0.417
54,543	7.75	25	PYK1	225.3	101.8	0.965
54,543	7.75	26	PYK1	39.8	101.8	0.965
55,221	6.66	146	YEL071W	16.3	3.0	0.244
55,295	4.35	134	PDI1	66.2	14.1	0.589
55,364	5.98	24	GLK1	22.6	6.0	0.237
55,481	7.97	118	ATP1	21.6	2.2	0.637
55,886	6.47	28	CYS4	22.2	NA <sup>c</sup>	0.444
56,167	5.83	33	ARO8	14.3	3.0	0.324
56,167	5.83	34	ARO8	9.1	3.0	0.324
56,584	6.36	20	CYB2	18.9	NA <sup>c</sup>	0.259
57,366	5.53	60	FRS2	2.3	0.7	0.451
57,383	5.98	144	ZWF1	5.6	0.7	0.215
57,464	5.49	36	THR4	21.4	3.7	0.508
57,512	5.50	7	SRV2	6.5	NA <sup>c</sup>	0.260
57,727	4.92	152	VMA2	33.7	8.9	0.546
58,573	6.47	17	ACH1	4.4	1.5	0.327
58,573	6.47	18	ACH1	5.4	1.5	0.327
61,353	5.87	21	PDC1	6.5	200.7	0.962
61,353	5.87	22	PDC1	303.2	200.7	0.962
61,353	5.87	23	PDC1	16.3	200.7	0.962
61,649	5.54	38	CCT8	2.2	1.5	0.271

Continued on following page

TABLE 1—Continued

Mol wt	pI	Spot no.	YPD gene name <sup>a</sup>	Protein abundance (10 <sup>3</sup> copies/cell)	mRNA abundance (copies/cell)	Codon bias
61,902	6.21	101	PDC5	4.2	NA <sup>c</sup>	0.828
62,266	6.19	16	ICL1	20.1	NA <sup>c</sup>	0.327
62,862	8.02	19	ILV3	5.3	4.5	0.548
63,082	6.40	119	PGM2	2.2	3.0	0.402
64,335	5.77	5	PAB1	30.4	1.5	0.616
66,120	5.42	8	STI1	6.7	0.7	0.313
66,120	5.42	9	STI1	6.4	0.7	0.313
66,450	5.29	141	SSB2	7.0	NA <sup>c</sup>	0.880
66,450	5.29	142	SSB2	2.3	NA <sup>c</sup>	0.880
66,456	5.23	10	SSB1	64.5	79.5	0.907
66,456	5.23	11	SSB1	59.0	79.5	0.907
66,456	5.23	12	SSB1	13.7	79.5	0.907
68,397	5.82	82	LEU4	3.1	3.0	0.407
69,313	4.90	13	SSA2	24.3	18.6	0.892
69,313	4.90	14	SSA2	77.1	18.6	0.892
74,378	8.46	15	YKL029C	2.8	3.7	0.353
75,396	5.82	6	GRS1	5.5	7.4	0.500
85,720	6.25	1	MET6	2.0	NA <sup>c</sup>	0.772
85,720	6.25	2	MET6	10.9	NA <sup>c</sup>	0.772
85,720	6.25	3	MET6	1.4	NA <sup>c</sup>	0.772
93,276	6.11	131	EFT1	17.9	41.6	0.890
93,276	6.11	132	EFT1	5.7	41.6	0.890
102,064 <sup>c</sup>	6.61 <sup>c</sup>	94	ADE3	4.8	5.2	0.423
107,482 <sup>c</sup>	5.33 <sup>c</sup>	4	MCM3	2.7	NA <sup>c</sup>	0.240

<sup>a</sup> YPD gene names are available from the YPD website (39).<sup>b</sup> NA, calculation could not be performed or was not available.<sup>c</sup> mRNA data inconclusive or NA.<sup>d</sup> No methionines in predicted ORF; therefore, protein concentration was not determined.<sup>e</sup> Measured molecular weight or pI did not match theoretical molecular weight or pI.

**Protein quantitation.** [<sup>35</sup>S]methionine-labeled gels were exposed to X-ray film overnight, and then the silver stain and film were used to excise 156 spots of varying intensities, molecular weights, and pIs. The excised spots were placed in 0.6-ml microcentrifuge tubes, and scintillation cocktail (100 µl) was added. The samples were vortexed and counted. In addition, two parallel gels were electroblotted to polyvinylidene difluoride membranes. The membranes were exposed to X-ray film, and four intense single spots were excised from each membrane and subjected to amino acid analysis. For these four spots, a mean of 209 ± 4 cpm/pmol of protein/methionine was found. This number was used to quantitate all remaining spots in conjunction with the number of methionines present in the proteins.

To ensure that proteins were labeled to equilibrium, parallel 2D gels were prepared and run on yeast metabolically labeled for 1, 2, 6, or 18 h. The corresponding 156 spots were excised from each gel, and radioactivity was measured by liquid scintillation counting for each spot. Calculated protein levels were highly reproducible for all time points measured after 1 h.

**Calculation of codon bias and predicted half-life.** Codon bias values were extracted from the YPD spreadsheet (17). Protein half-lives were calculated based on the N-end rule (33). When the N-terminal processing was not known experimentally, it was predicted based on the affinity of methionine aminopeptidase (31).

## RESULTS

**Characteristics of proteome approach.** Nearly every facet of proteome analysis hinges on the unambiguous identification of large numbers of expressed proteins in cells. Several techniques have been described previously for the identification of proteins separated by 2DE, including N-terminal and internal sequencing (1, 2), amino acid analysis (38), and more recently mass spectrometry (25). We utilized techniques based on mass spectrometry because they afford the highest levels of sensitivity and provide unambiguous identification. The specific procedure used is schematically illustrated in Fig. 1 and is based on three principles. First, proteins are removed from the gel by

proteolytic in-gel digestion, and the resulting peptides are separated by on-line capillary high-performance liquid chromatography. Second, the eluting peptides are ionized and detected, and the specific peptide ions are selected and fragmented by the mass spectrometer. To achieve this, the mass spectrometer switches between the MS mode (for peptide mass identification) and the MS/MS mode (for peptide characterization and sequencing). Selected peptides are fragmented by a process called collision-induced dissociation (CID) to generate a tandem mass spectrum (MS/MS spectrum) that contains the peptide sequence information. Third, individual CID mass spectra are then compared by computer algorithms to predicted spectra from a sequence database. This results in the identification of the peptide and, by association, the protein(s) in the spot. Unambiguous protein identification is attained in a single analysis by the detection of multiple peptides derived from the same protein.

**Protein identification.** Yeast total cell protein lysate (40 µg), metabolically labeled with [<sup>35</sup>S]methionine, was electrophoretically separated by isoelectric focusing in the first dimension and by SDS-10% polyacrylamide gel electrophoresis in the second dimension. Proteins were visualized by silver staining and by autoradiography. Of the more than 1,000 proteins visible by silver staining, 156 spots were excised from the gel and subjected to in-gel tryptic digestion, and the resulting peptides were analyzed and identified by microspray LC-MS/MS techniques as described above. The proteins in this study were all identified automatically by computer software with no human interpretation of mass spectra. They are indicated in Fig. 2 and detailed in Table 1.

The CID spectra shown in Fig. 3 indicate that the quality of the identification data generated was suitable for unambiguous protein identification. The spectra represent the amino acid sequences of tryptic peptides NSGDIVNLGSLAGR (Fig. 3A) and FAVGAFTDSLRL (Fig. 3B). Both peptides were derived from protein S57593 (hypothetical protein YMR226C), which migrated to spot 114 (molecular weight, 29,156; pI, 6.59) in the 2D gel in Fig. 2. Five other peptides from the same analysis were also computer matched to the same protein sequence.

**Protein and mRNA quantitation.** For the 156 genes investigated, the protein expression levels ranged from 2,200 (PGM2) to 863,000 (TDH2/TDH3) copies/cell. The levels of mRNA for each of the genes identified were calculated from SAGE frequency tables (35). These tables contain the mRNA levels for 4,665 genes in yeast strain YPH499 grown to mid-log phase in YPD medium on glucose as a carbon source. In some instances, the mRNA levels could not be calculated for reasons stated in Materials and Methods. For the proteins analyzed in this study, mean transcript levels varied from 0.7 to 473 copies/cell.

**Selection of the sample population for mRNA-protein expression level correlation.** The protein spots selected for identification were selected from spots visible by silver staining in the 2D gel. An attempt was made not to include spots where overlap with other spots was readily apparent. The number of proteins identified was 156 (Table 1). Some proteins migrated to more than one spot (presumably due to differential protein processing or modifications), and protein levels from these spots were calculated by integrating the intensities of the different spots. The 156 protein spots analyzed represented the products of 128 different genes. Genes were excluded from the correlation analysis only if part of the data set was missing; i.e., genes were excluded if (i) no mRNA expression data were available for the protein or putative SAGE tags were ambiguous, (ii) the amino acid sequence did not contain methionine, (iii) more than a single protein was conclusively identified as

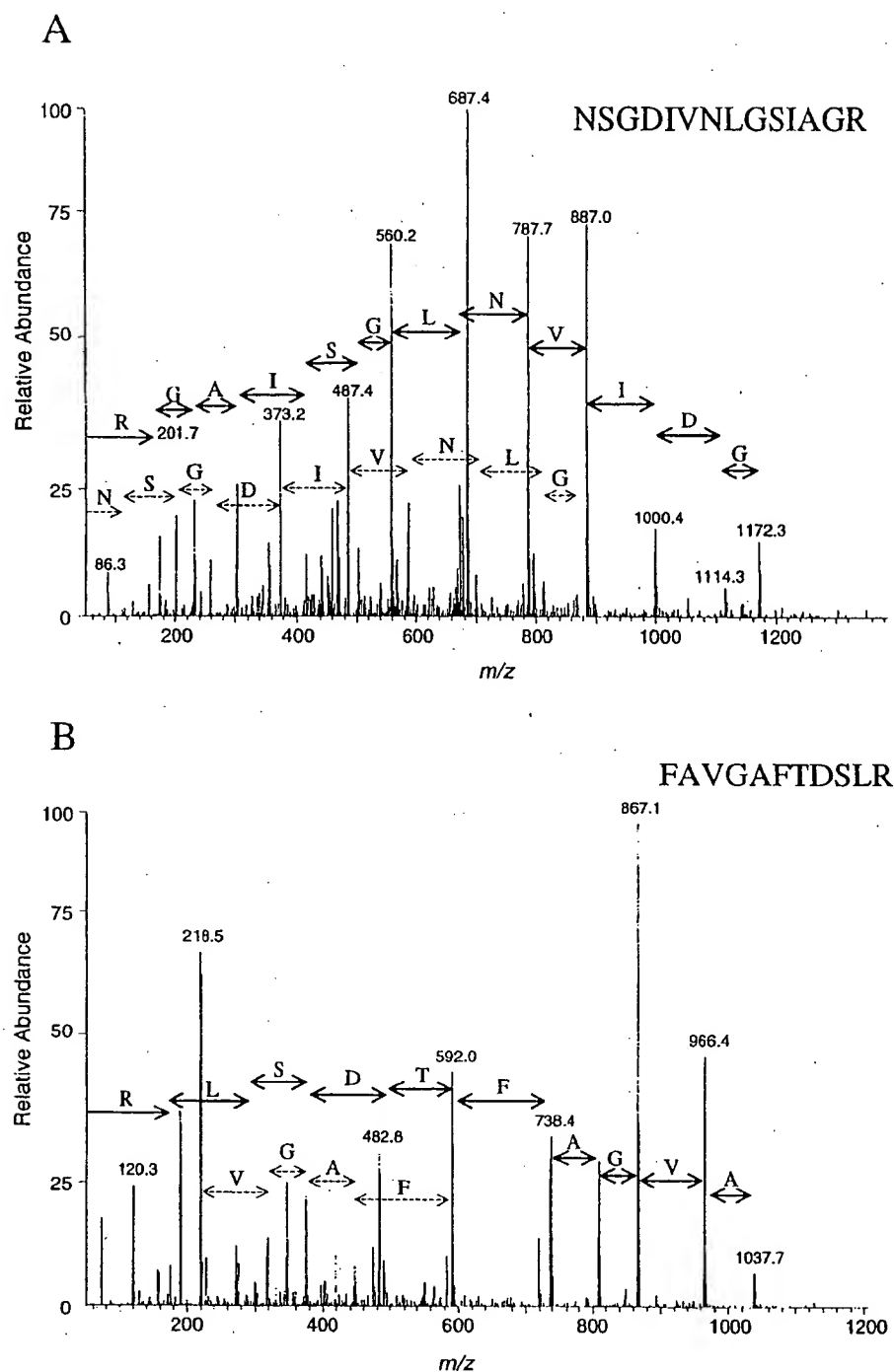


FIG. 3. Tandem mass (MS/MS) spectra resulting from analysis of a single spot on a 2D gel. The first quadrupole selected a single mass-to-charge ratio ( $m/z$ ) of 687.2 (A) or 592.6 (B), while the collision cell was filled with argon gas, and a voltage which caused the peptide to undergo fragmentation by CID was applied. The third quadrupole scanned the mass range from 50 to 1,400  $m/z$ . The computer program Sequest (8) was utilized to match MS/MS spectra to amino acid sequence by database searching. Both spectra matched peptides from the same protein, S57593 (yeast hypothetical protein YMR226C). Five other peptides from the same analysis were matched to the same protein.

migrating to the same gel spot, or (iv) the theoretical and observed pIs and molecular weights could not be reconciled. After these criteria were applied, the number of genes used in the correlation analysis was 106.

**Codon bias and predicted half-lives.** Codon bias is thought to be an indicator of protein expression, with highly expressed proteins having large codon bias values. The codon bias distribution for the entire set of more than 6,000 predicted yeast

gene ORFs is presented in Fig. 4A. The interval with the largest frequency of genes is between the codon bias values of 0.0 and 0.1. This segment contains more than 2,500 genes. The distribution of the codon bias values of the 128 different genes found in this study (all protein spots from Fig. 2) is shown in Fig. 4B, and protein half-lives (predicted from applying the N-end rule [33] to the experimentally determined or predicted protein N termini) are shown in Fig. 4C. No genes were identified with codon bias values less than 0.1 even though thousands of genes exist in this category. In addition, nearly all of the proteins identified had long predicted half-lives (greater than 30 h).

**Correlation of mRNA and protein expression levels.** The correlation between mRNA and protein levels of the genes selected as described above is shown in Fig. 5. For the entire group (106 genes) for which a complete data set was generated, there was a general trend of increased protein levels resulting from increased mRNA levels. The Pearson product moment correlation coefficient for the whole data set (106 genes) was 0.935. This number is highly biased by a small number of genes with very large protein and message levels. A more representative subset of the data is shown in the inset of Fig. 5. It shows genes for which the message level was below 10 copies/cell and includes 69% (73 of 106 genes) of the data used in the study. The Pearson product moment correlation coefficient for this data set was only 0.356. We also found that levels of protein expression coded for by mRNA with comparable abundance varied by as much as 30-fold and that the mRNA levels coding for proteins with comparable expression levels varied by as much as 20-fold.

The distortion of the correlation value induced by the uneven distribution of the data points along the x axis is further demonstrated by the analysis in Fig. 6. The 106 samples included in the study were ranked by protein abundance, and the Pearson product moment correlation coefficient was repeatedly calculated after including progressively more, and higher-abundance, proteins in each calculation. The correlation values remained relatively stable in the range of 0.1 to 0.4 if the lowest-expressed 40 to 95 proteins used in this study were included. However, the correlation value steadily climbed by the inclusion of each of the 11 very highly expressed proteins.

**Correlation of protein and mRNA expression levels with codon bias.** Codon bias is the propensity for a gene to utilize the same codon to encode an amino acid even though other codons would insert the identical amino acid in the growing polypeptide sequence. It is further thought that highly expressed proteins have large codon biases (3). To assess the value of codon bias for predicting mRNA and protein levels in exponentially growing yeast cells, we plotted the two experimental sets of data versus the codon bias (Fig. 7). The distribution patterns for both mRNA and protein levels with respect to codon bias were highly similar. There was high variability in the data within the codon bias range of 0.8 to 1.0. Although a large codon bias generally resulted in higher protein and message expression levels, codon bias did not appear to be predictive of either protein levels or mRNA levels in the cell.

## DISCUSSION

The desired end point for the description of a biological system is not the analysis of mRNA transcript levels alone but also the accurate measurement of protein expression levels and their respective activities. Quantitative analysis of global mRNA levels currently is a preferred method for the analysis of the state of cells and tissues (11). Several methods which either provide absolute mRNA abundance (34, 35) or relative

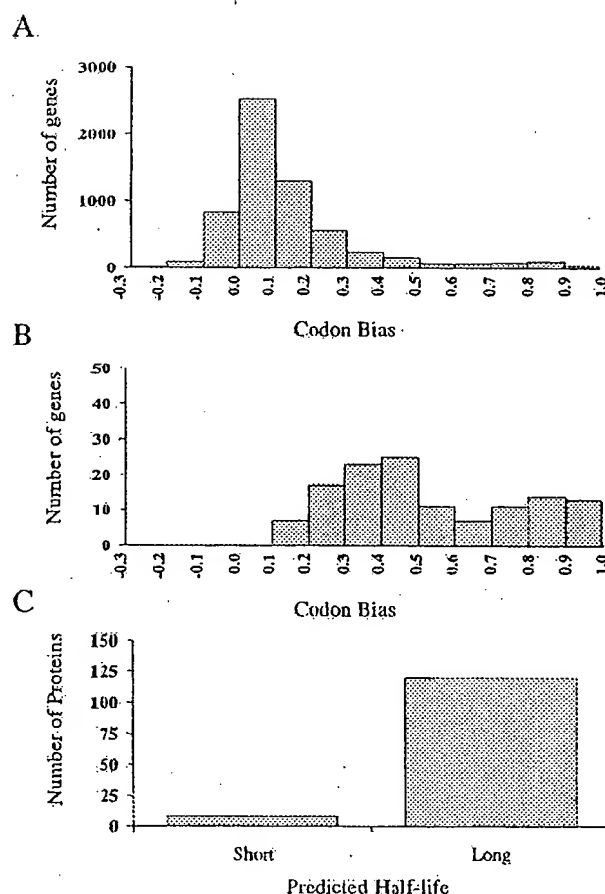


FIG. 4. Current proteome analysis technology utilizing 2DE without pre-enrichment samples mainly highly expressed and long-lived proteins. Genes encoding highly expressed proteins generally have large codon bias values. (A) Distribution of the yeast genome (more than 6,000 genes) based on codon bias. The interval with the largest frequency of genes is 0.0 to 0.1, with more than 2,500 genes. (B) Distribution of the genes from identified proteins in this study based on codon bias. No genes with codon bias values less than 0.1 were detected in this study. (C) Distribution of identified proteins in this study based on predicted half-life (estimated by N-end rule).

mRNA levels in comparative analyses (20, 27) have been described elsewhere. The techniques are fast and exquisitely sensitive and can provide mRNA abundance for potentially any expressed gene. Measured mRNA levels are often implicitly or explicitly extrapolated to indicate the levels of activity of the corresponding protein in the cell. Quantitative analysis of protein expression levels (proteome analysis) is much more time-consuming because proteins are analyzed sequentially one by one and is not general because analyses are limited to the relatively highly expressed proteins. Proteome analysis does, however, provide types of data that are of critical importance for the description of the state of a biological system and that are not readily apparent from the sequence and the level of expression of the mRNA transcript. This study attempts to examine the relationship between mRNA and protein expression levels for a large number of expressed genes in cells representing the same state.

Limits in the sensitivity of current protein analysis technology precluded a completely random sampling of yeast proteins. We therefore based the study on those proteins visible by silver

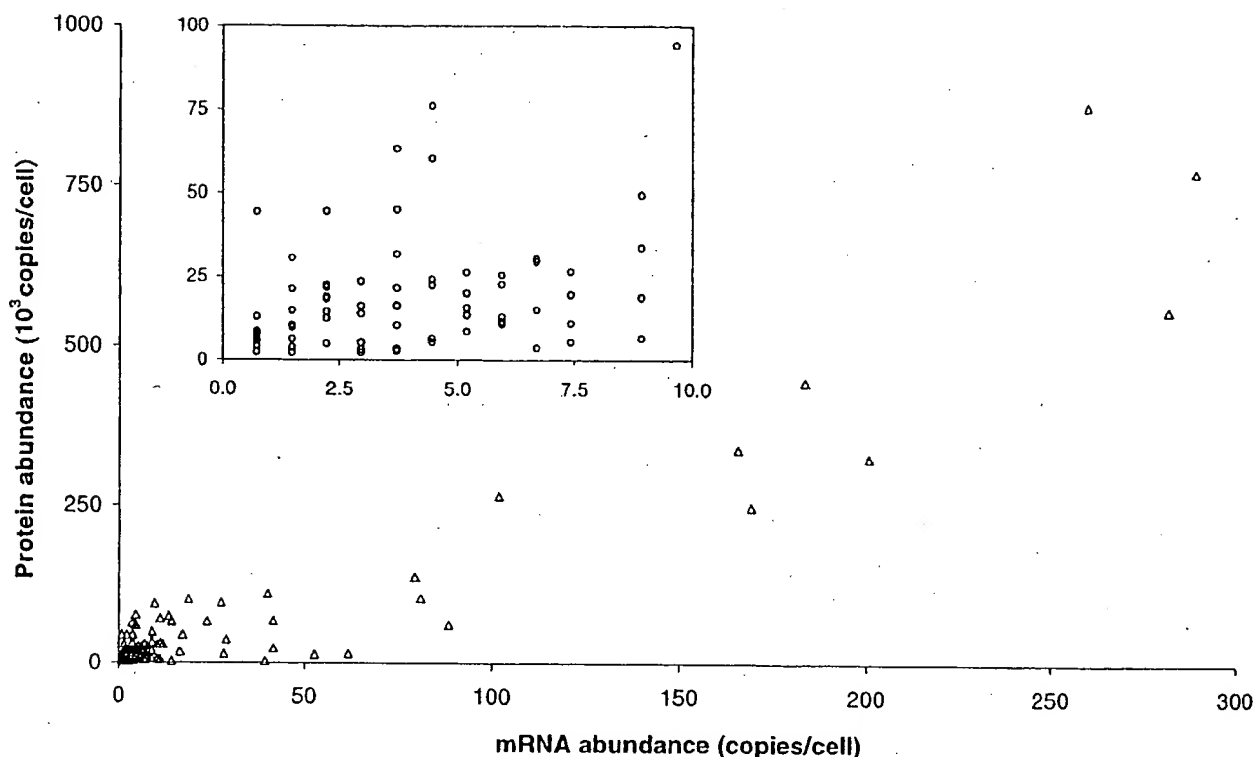


FIG. 5. Correlation between protein and mRNA levels for 106 genes in yeast growing at log phase with glucose as a carbon source. mRNA and protein levels were calculated as described in Materials and Methods. The data represent a population of genes with protein expression levels visible by silver staining on a 2D gel chosen to include the entire range of molecular weights, isoelectric focusing points, and staining intensities. The inset shows the low-end portion of the main figure. It contains 69% of the original data set. The Pearson product moment correlation for the entire data set was 0.935. The correlation for the inset containing 73 proteins (69%) was only 0.356.

staining on a 2D gel. Of the more than 1,000 visible spots, 156 were chosen to include the entire range of molecular weights, isoelectric focusing points, and staining intensities displayed on the 2D protein pattern. The genes identified in this study shared a number of properties. First, all of the proteins in this study had a codon bias of greater than 0.1 and 93% were greater than 0.2 (Fig. 4B). Second, with few exceptions, the proteins in this study had long predicted half-lives according to the N-end rule (Fig. 4C). Third, low-abundance proteins with regulatory functions such as transcription factors or protein kinases were not identified.

Because the population of proteins used in this study appears to be fairly homogeneous with respect to predicted half-life and codon bias, it might be expected that the correlation of the mRNA and protein expression levels would be stronger for this population than for a random sample of yeast proteins. We tested this assumption by evaluating the correlation value if different subsets of the available data were included in the calculation. The 106 proteins were ranked from lowest to highest protein expression level, and the trend in the correlation value was evaluated by progressively including more of the higher-abundance proteins in the calculation (Fig. 6). The correlation value when only the lower-abundance 40 to 93 proteins were examined was consistently between 0.1 and 0.4. If the 11 most abundant proteins were included, the correlation steadily increased to 0.94. We therefore expect that the correlation for all yeast proteins or for a random selection would be less than 0.4. The observed level of correlation between mRNA and protein expression levels suggests the importance

of posttranslational mechanisms controlling gene expression. Such mechanisms include translational control (15) and control of protein half-life (33). Since these mechanisms are also active in higher eukaryotic cells, we speculate that there is no predictive correlation between steady-state levels of mRNA and those of protein in mammalian cells.

Like other large-scale analyses, the present study has several potential sources of error related to the methods used to determine mRNA and protein expression levels. The mRNA levels were calculated from frequency tables of SAGE data. This method is highly quantitative because it is based on actual sequencing of unique tags from each gene, and the number of times that a tag is represented is proportional to the number of mRNA molecules for a specific gene. This method has some limitations including the following: (i) the magnitude of the error in the measurement of mRNA levels is inversely proportional to the mRNA levels, (ii) SAGE tags from highly similar genes may not be distinguished and therefore are summed, (iii) some SAGE tags are from sequences in the 3' untranslated region of the transcript, (iv) incomplete cleavage at the SAGE tag site by the restriction enzyme can result in two tags representing one mRNA, and (v) some transcripts actually do not generate a SAGE tag (34, 35).

For the SAGE method, the error associated with a value increases with a decreasing number of transcripts per cell. The conclusions drawn from this study are dependent on the quality of the mRNA levels from previously published data (35). Since more than 65% of the mRNA levels included in this study were calculated to 10 copies/cell or less (40% were less

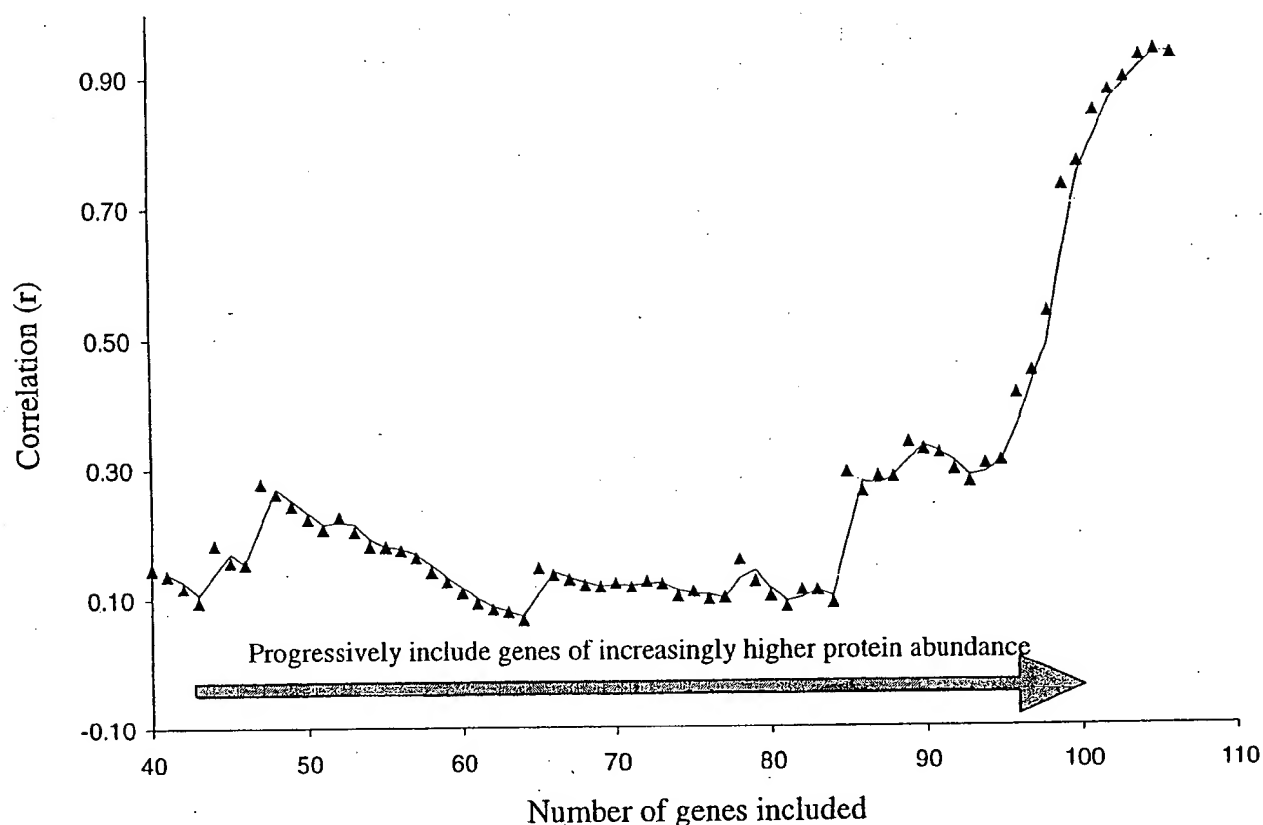


FIG. 6. Effect of highly abundant proteins on Pearson product moment correlation coefficient for mRNA and protein abundance in yeast. The set of 106 genes was ranked according to protein abundance, and the correlation value was calculated by including the 40 lowest-abundance genes and then progressively including the remaining 66 genes in order of abundance. The correlation value climbs as the final 11 highly abundant proteins are included.

than 4 copies/cell), the error associated with these values may be quite large. The mRNA levels were calculated from more than 20,000 transcripts. Assuming that the estimate of 15,000 mRNA molecules per cell is correct (16), this would mean that mRNA transcripts present at only a single copy per cell would be detected 72% of the time (35). The mRNA levels for each gene were carefully scrutinized, and only mRNA levels for which a high degree of confidence existed were included in the correlation value.

Protein abundance was determined by metabolic radiolabeling with [ $^{35}\text{S}$ ]methionine. The calculation required knowledge of three variables: the number of methionines in the mature protein, the radioactivity contained in the protein, and the specific activity of the radiolabel normalized per methionine. The number of methionines per protein was determined from the amino acid sequence of the proteins identified by tandem mass spectrometry. For some proteins, it was not known whether the methionine of the nascent polypeptide was processed away. The N termini of those proteins were predicted based on the specificity of methionine aminopeptidase (31). If the N-terminal processing did not conform to the predicted specificity of processing enzymes, the calculation of the number of methionines would be affected. This discrepancy would affect most the quantitation of a protein with a very low number of methionines. The average number of calculated methionines per protein in this study was 7.2. We therefore expect the potential for erroneous protein quantitation due to unusual N-terminal processing to be small.

The amount of radioactivity contained in a single spot might be the sum of the radioactivity of comigrating proteins. Because protein identification was based on tandem mass spectrometric techniques, comigrating proteins could be identified. However, comigrating proteins were rarely detected in this study, most likely because relatively small amounts of total protein (40  $\mu\text{g}$ ) were initially loaded onto the gels, which resulted in highly focused spots containing generally 1 to 25 ng of protein. Because of the relatively small amount loaded, the concentrations of any potentially comigrating protein would likely be below the limit of detection of the mass spectrometry technique used in this study (1 to 5 ng) and below the limit of visualization by silver staining (1 to 5 ng). In the overwhelming majority of the samples analyzed, numerous peptides from a single protein were detected. It is assumed that any comigrating proteins were at levels too low to be detected and that their influence in the calculation would be small.

The specific activity of the radiolabel was determined by relating the precise amount of protein present in selected spots of a parallel gel, as determined by quantitative amino acid composition analysis, to the number of methionines present in the sequence of those proteins and the radioactivity determined by liquid scintillation counting. It is possible that the resulting number might be influenced by unavoidable losses inherent in the amino acid analysis procedure applied. Because four different proteins were utilized in the calculation and the experiment was done in duplicate, the specific activity calculated is thought to be highly accurate. Indeed, the specific



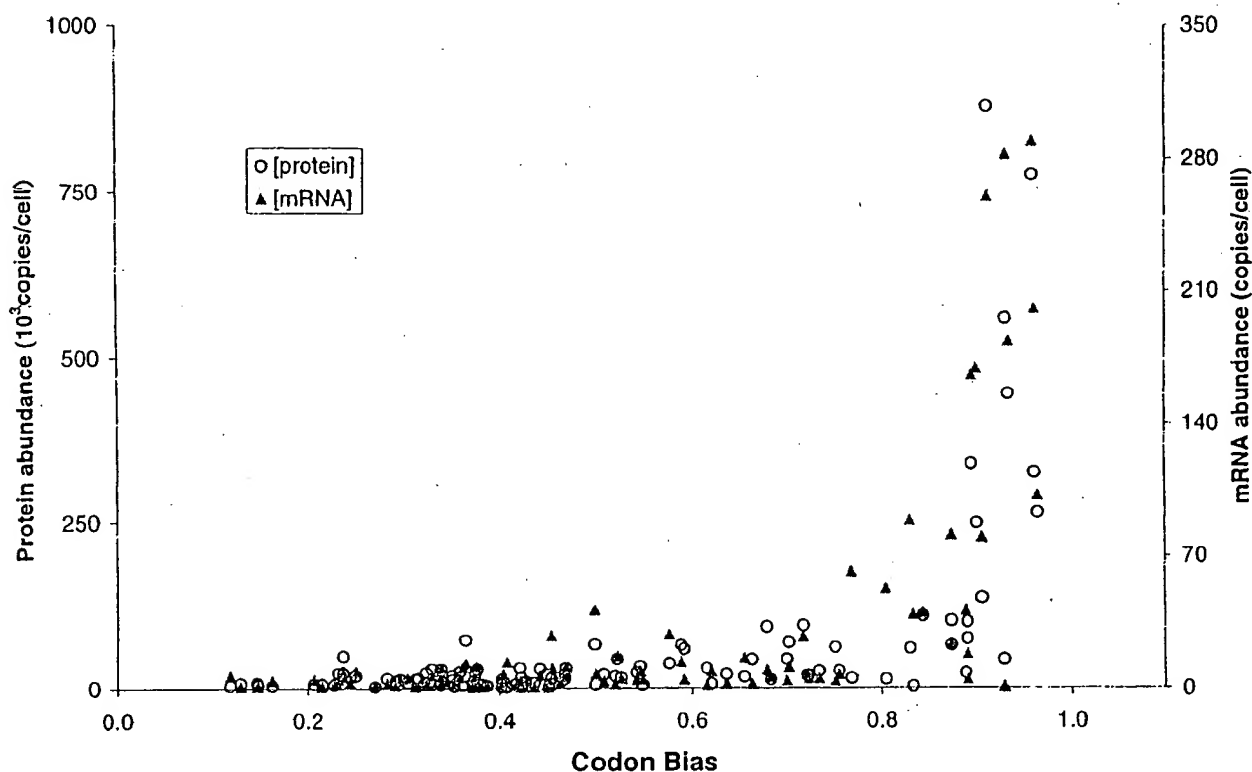


FIG. 7. Relationship between codon bias and protein and mRNA levels in this study. Yeast mRNA and protein expression levels were calculated as described in Materials and Methods. The data represent the same 106 genes as in Fig. 5.

activities calculated for each of the four proteins varied by less than 10%. Any inconsistencies in the calculation of the specific activity would result in differences in the absolute levels calculated but not in the relative numbers and would therefore not influence the correlation value determined.

The protein quantitative method used eliminates a number of potential errors inherent in previous methods for the quantitation of proteins separated by 2DE, such as preferential protein staining and bias caused by inequalities in the number of radiolabeled residues per protein. Any 2D gel-based method of quantitation is complicated by the fact that in some cases the translation products of the same mRNA migrated to different spots. One major reason is posttranslational modification or processing of the protein. Also, artifactual proteolysis during cell lysis and sample preparation can lead to multiple resolved forms of the protein. In such cases, the protein levels of spots coded for by the same mRNA were pooled. In addition, the existence of other spots coded for by the same mRNA that were not analyzed by mass spectrometry or that were below the limit of detection for silver staining cannot be ruled out. However, since this study is based on a class of highly expressed proteins, the presence of undetected minor spots below silver staining sensitivity corresponding to a protein analyzed in the study would generally cause a relatively small error in protein quantitation.

Codon bias is a measure of the propensity of an organism to selectively utilize certain codons which result in the incorporation of the same amino acid residue in a growing polypeptide chain. There are 61 possible codons that code for 20 amino acids. The larger the codon bias value, the smaller the number of codons that are used to encode the protein (19). It is

thought that codon bias is a measure of protein abundance because highly expressed proteins generally have large codon bias values (3, 13).

Nearly all of the most highly expressed proteins had codon bias values of greater than 0.8. However, we detected a number of genes with high codon bias and relative low protein abundance (Fig. 7). For example, the expressed gene with both the second largest protein and mRNA levels in the study was *ENO2\_YEAST* (775,000 and 289.1 copies/cell, respectively). *ENO1\_YEAST* was also present in the gel at much lower protein and mRNA levels (44,200 and 0.7 copies/cell, respectively). The codon bias values for *ENO2* and *ENO1* are similar (0.96 and 0.93, respectively), but the expression of the two genes is differentially regulated. Specifically, *ENO1\_YEAST* is glucose repressed (6) and was therefore present in low abundance under the conditions used. Other genes with large codon bias values that were not of high protein abundance in the gel include *EFT1*, *TIF1*, *HXK2*, *GSP1*, *EGD2*, *SHM2*, and *TAL1*. We conclude that merely determining the codon bias of a gene is not sufficient to predict its protein expression level.

Interestingly, codon bias appears to be an excellent indicator of the boundaries of current 2D gel proteome analysis technology. There are thousands of genes with expressed mRNA and likely expressed protein with codon bias values less than 0.1 (Fig. 4A). In this study, we detected none of them, and only a very small percentage of the genes detected in this study had codon bias values between 0.1 and 0.2 (Fig. 4B). Indeed, in every examined yeast proteome study (5, 7, 13, 28) where the combined total number of identified proteins is 300 to 400, this same observation is true. It is expected that for the more complex cells of higher eukaryotic organisms the detection of



low-abundance proteins would be even more challenging than for yeast. This indicates that highly abundant, long-lived proteins are overwhelmingly detected in proteome studies. If proteome analysis is to provide truly meaningful information about cellular processes, it must be able to penetrate to the level of regulatory proteins, including transcription factors and protein kinases. A promising approach is the use of narrow-range focusing gels with immobilized pH gradients (IPG) (23). This would allow for the loading of significantly more protein per pH unit covered and also provide increased resolution of proteins with similar electrophoretic mobilities. A standard pH gradient in an isoelectric focusing gel covers a 7-pH-unit range (pH 3 to 10) over 18 cm. A narrow-range focusing gel might expand the range to 0.5 pH units over 18 cm or more. This could potentially increase by more than 10-fold the number of proteins that can be detected. Clearly, current proteome technology is incapable of analyzing low-abundance regulatory proteins without employing an enrichment method for relatively low-abundance proteins. In conclusion, this study examined the relationship between yeast protein and message levels and revealed that transcript levels provide little predictive value with respect to the extent of protein expression.

#### ACKNOWLEDGMENTS

This work was supported by the National Science Foundation Science and Technology Center for Molecular Biotechnology, NIH grant T32HG00035-3, and a grant from Oxford Glycosciences.

We thank Jimmy Eng for expert computer programming, Garry Corthals and John R. Yates III for critical discussion, and Siavash Mohandesi for expert technical help.

#### REFERENCES

- Aebersold, R. H., J. Leavitt, R. A. Saavedra, L. E. Hood, and S. B. Kent. 1987. Internal amino acid sequence analysis of proteins separated by one- or two-dimensional gel electrophoresis after *in situ* protease digestion on nitrocellulose. *Proc. Natl. Acad. Sci. USA* 84:6970-6974.
- Aebersold, R. H., D. B. Teplow, L. E. Hood, and S. B. Kent. 1986. Electrophoretic blotting onto activated glass. High efficiency preparation of proteins from analytical sodium dodecyl sulfate-polyacrylamide gels for direct sequence analysis. *Eur. J. Biochem.* 261:4229-4238.
- Bennetzen, J. L., and B. D. Hall. 1982. Codon selection in yeast. *J. Biol. Chem.* 257:3026-3031.
- Boucherie, H., G. Dujardin, M. Kermorgant, C. Mourillot, P. Slonimski, and M. Perrot. 1995. Two-dimensional protein map of *Saccharomyces cerevisiae*: construction of a gene-protein index. *Yeast* 11:601-613.
- Boucherie, H., F. Sagliocco, R. Joubert, I. Maillet, J. Labarre, and M. Perrot. 1996. Two-dimensional gel protein database of *Saccharomyces cerevisiae*. *Electrophoresis* 17:1683-1699.
- Carmen, A. A., P. K. Brindle, C. S. Park, and M. J. Holland. 1995. Transcriptional regulation by an upstream repression sequence from the yeast enolase gene ENO1. *Yeast* 11:1031-1043.
- Ducet, A., I. VanOostveen, J. K. Eng, J. R. Yates, and R. Aebersold. 1998. High throughput protein characterization by automated reverse-phase chromatography/electrospray tandem mass spectrometry. *Protein Sci.* 7:706-719.
- Eng, J., A. McCormack, and J. R. Yates. 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* 5:976-989.
- Figgs, D., A. Ducet, J. R. Yates, and R. Aebersold. 1996. Protein identification by solid phase microextraction-capillary zone electrophoresis-microelectrospray-tandem mass spectrometry. *Nat. Biotechnol.* 14:1579-1583.
- Figgs, D., I. VanOostveen, A. Ducet, and R. Aebersold. 1996. Protein identification by capillary zone electrophoresis/microelectrospray ionization-tandem mass spectrometry at the subfemtomole level. *Anal. Chem.* 68:1822-1828.
- Fraser, C. M., and R. D. Fleischmann. 1997. Strategies for whole microbial genome sequencing and analysis. *Electrophoresis* 18:1207-1216.
- Garrels, J. I., B. Futcher, R. Kobayashi, G. I. Latter, B. Schwender, T. Volpe, J. R. Warner, and C. S. McLaughlin. 1994. Protein identifications for a *Saccharomyces cerevisiae* protein database. *Electrophoresis* 15:1466-1486.
- Garrels, J. I., C. S. McLaughlin, J. R. Warner, B. Futcher, G. I. Latter, R. Kobayashi, B. Schwender, T. Volpe, D. S. Anderson, F. Mesquita-Fuentes, and W. E. Payne. 1997. Proteome studies of *Saccharomyces cerevisiae*: identification and characterization of abundant proteins. *Electrophoresis* 18:1347-1360.
- Gygi, S. P., and R. Aebersold. 1998. Absolute quantitation of 2-DE protein spots, p. 417-421. In A. J. Link (ed.), 2-D protocols for proteome analysis. Humana Press, Totowa, N.J.
- Harford, J. B., and D. R. Morris. 1997. Post-transcriptional gene regulation. Wiley-Liss, Inc., New York, N.Y.
- Hereford, L. M., and M. Rosbash. 1977. Number and distribution of polyadenylated RNA sequences in yeast. *Cell* 10:453-462.
- Hodges, P. E., W. E. Payne, and J. L. Garrels. 1998. The Yeast Protein Database (YPD): a curated proteome database for *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 26:68-72.
- Klose, J., and U. Kobalz. 1995. Two-dimensional electrophoresis of proteins: an updated protocol and implications for a functional analysis of the genome. *Electrophoresis* 16:1034-1059.
- Kuriland, C. G. 1991. Codon bias and gene expression. *FEBS Lett.* 285:165-169.
- Lashkari, D. A., J. L. DeRisi, J. H. McCusker, A. P. Namath, C. Gentile, S. Y. Hwang, P. O. Brown, and R. W. Davis. 1997. Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc. Natl. Acad. Sci. USA* 94:13057-13062.
- Liang, P., and A. B. Pardee. 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 257:967-971.
- Link, A. J., L. G. Hays, E. B. Carmack, and J. R. Yates III. 1997. Identifying the major proteome components of *Haemophilus influenzae* type-strain NCTC 8143. *Electrophoresis* 18:1314-1334.
- Nawrocki, A., M. R. Larsen, A. V. Podtelejnikov, O. N. Jensen, M. Mann, P. Roepstorff, A. Gorg, S. J. Fey, and P. M. Larsen. 1998. Correlation of acidic and basic carrier ampholyte and immobilized pH gradient two-dimensional gel electrophoresis patterns based on mass spectrometric protein identification. *Electrophoresis* 19:1024-1035.
- O'Farrell, P. H. 1975. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* 250:4097-4021.
- OWL Protein Sequence Database. 2 August 1998, posting date. [Online.] <http://bmb5g11.leeds.ac.uk/bmb5dp/owl.html>. [8 January 1999, last date accessed.]
- Patterson, S. D., and R. Aebersold. 1995. Mass spectrometric approaches for the identification of gel-separated proteins. *Electrophoresis* 16:1791-1814.
- Pennington, S. R., M. R. Wilkins, D. F. Hochstrasser, and M. J. Dunn. 1997. Proteome analysis: from protein characterization to biological function. *Trends Cell Biol.* 7:168-173.
- Shalon, D., S. J. Smith, and P. O. Brown. 1996. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res.* 6:639-645.
- Shevchenko, A., O. N. Jensen, A. V. Podtelejnikov, F. Sagliocco, M. Wilm, O. Vorm, P. Mortensen, H. Boucherie, and M. Mann. 1996. Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc. Natl. Acad. Sci. USA* 93:14440-14445.
- Shevchenko, A., M. Wilm, O. Vorm, and M. Mann. 1996. Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Anal. Chem.* 68:850-858.
- Sikorski, R. S., and P. Hieter. 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* 122:19-27.
- Tsumasawa, S., J. W. Stewart, and F. Sherman. 1985. Amino-terminal processing of mutant forms of yeast iso-1-cytochrome c. The specificities of methionine aminopeptidase and acetyltransferase. *J. Biol. Chem.* 260:5382-5391.
- Ullinger, S., K. Kuchler, T. H. Meyer, S. Uebel, and R. Tampé. 1997. Intracellular location, complex formation, and function of the transporter associated with antigen processing in yeast. *Eur. J. Biochem.* 245:266-272.
- Varshavsky, A. 1996. The N-end rule: functions, mysteries, uses. *Proc. Natl. Acad. Sci. USA* 93:12142-12149.
- Velculescu, V. E., L. Zhang, B. Vogelstein, and K. W. Kinzler. 1995. Serial analysis of gene expression. *Science* 270:484-487.
- Velculescu, V. E., L. Zhang, W. Zhou, J. Vogelstein, M. A. Basrai, D. E. Bassett, Jr., P. Hieter, B. Vogelstein, and K. W. Kinzler. 1997. Characterization of the yeast transcriptome. *Cell* 88:243-251.
- Wilkins, M. R., K. L. Williams, R. D. Appel, and D. F. Hochstrasser. 1997. Proteome research: new frontiers in functional genomics. Springer-Verlag, Berlin, Germany.
- Wilm, M., A. Shevchenko, T. Houthaeve, S. Breit, L. Schweigerer, T. Fotsis, and M. Mann. 1996. Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* 379:466-469.
- Yan, J. X., M. R. Wilkins, K. On, A. A. Goolley, K. L. Williams, J. C. Sanchez, O. Golaz, C. Pasquali, and D. F. Hochstrasser. 1996. Large-scale amino-acid analysis for proteome studies. *J. Chromatogr. A* 736:291-302.
- YPD Website. 6 March 1998, revision date. [Online.] Proteome, Inc. <http://www.proteome.com/YPDhome.html>. [8 January 1999, last date accessed.]



## ORIGINAL ARTICLE

# Identification of putative oncogenes in lung adenocarcinoma by a comprehensive functional genomic approach

R Li<sup>1</sup>, H Wang<sup>1</sup>, BN Bekele<sup>2</sup>, Z Yin<sup>3</sup>, NP Caraway<sup>1</sup>, RL Katz<sup>1</sup>, SA Stass<sup>4</sup> and F Jiang<sup>1</sup>

<sup>1</sup>Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; <sup>2</sup>Department of Biostatistics & Applied Mathematics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; <sup>3</sup>Department of Internal Medicine, The University of Texas Houston Medical School, Houston, TX, USA and <sup>4</sup>Department of Pathology, The University of Maryland School of Medicine, Baltimore, MD, USA

Amplification and overexpression of putative oncogenes confer growth advantages for tumor development. We used a functional genomic approach that integrated simultaneous genomic and transcript microarray, proteomics, and tissue microarray analyses to directly identify putative oncogenes in lung adenocarcinoma. We first identified 183 genes with increases in both genomic copy number and transcript in six lung adenocarcinoma cell lines. Next, we used two-dimensional polyacrylamide gel electrophoresis and mass spectrometry to identify 42 proteins that were overexpressed in the cancer cells relative to normal cells. Comparing the 183 genes with the 42 proteins, we identified four genes – *PRDX1*, *EEF1A2*, *CALR*, and *KCIP-1* – in which elevated protein expression correlated with both increased DNA copy number and increased transcript levels (all  $r > 0.84$ , two-sided  $P < 0.05$ ). These findings were validated by Southern, Northern, and Western blotting. Specific inhibition of *EEF1A2* and *KCIP-1* expression with siRNA in the four cell lines tested suppressed proliferation and induced apoptosis. Parallel fluorescence *in situ* hybridization and immunohistochemical analyses of *EEF1A2* and *KCIP-1* in tissue microarrays from patients with lung adenocarcinoma showed that gene amplification was associated with high protein expression for both genes and that protein overexpression was related to tumor grade, disease stage, Ki-67 expression, and a shorter survival of patients. The amplification of *EEF1A2* and *KCIP-1* and the presence of overexpressed protein in tumor samples strongly suggest that these genes could be oncogenes and hence potential targets for diagnosis and therapy in lung adenocarcinoma. *Oncogene* (2006) 25, 2628–2635. doi:10.1038/sj.onc.1209289; published online 12 December 2005

**Keywords:** lung cancer; microarrays; proteomics; tissue microarray

## Introduction

In lung adenocarcinoma, as in other types of cancer, gene amplification and the consequent overexpression of the amplified oncogene play an important role in the development of tumors, because their overexpression confers a growth advantage. The ability to identify putative oncogenes that are activated during tumorigenesis could facilitate the choice of molecular genetic targets for diagnosis and therapy of the disease. This concept has been exemplified by *HER-2*, which was first found to be amplified in neuroblastomas and subsequently shown to be associated with poor prognosis in breast cancer (Ross and Fletcher, 1999). Now, *HER-2* aberrations are used as a predictor of response to therapy, and treatment of *HER-2*-positive breast cancer with the monoclonal anti-*HER-2* antibody trastuzumab has been shown to improve prognosis (Ross and Fletcher, 1999). Emerging evidence of common amplicons in lung adenocarcinomas (Luk *et al.*, 2001; Jiang *et al.*, 2004; Tonon *et al.*, 2005) suggests that additional oncogenes remain to be identified; however, conventional techniques are ineffective in pinpointing such oncogenes. Parallel measurement of DNA copy number and mRNA levels in cDNA microarrays permits changes in copy number to be compared with transcription levels on a gene-by-gene basis to generate lists of candidate genes within the defining amplicons (Hyman *et al.*, 2002; Pollack *et al.*, 2002). However, use of transcript patterns does not allow assessment of the expression of protein products or identification of proto-oncogenes. Another approach, identifying differentially expressed proteins by proteomic analysis and then comparing the proteins present with mRNA expression in cDNA microarrays from the same specimens, can clarify the extent to which changes in transcript patterns reflect changes in their cognate proteins and post-transcriptional mechanisms (Chen *et al.*, 2002), but this approach cannot be used to identify oncogenes driven by extensive increases of their gene copy number. Moreover, using individual microarrays or proteomic approaches alone cannot distinguish the cancer-driving oncogenes that directly propel tumor progression from the larger number of passenger genes that may be concurrently over-represented but are not biologically relevant in tumor development.

Correspondence: Assistant Professor F Jiang, Department of Pathology, The University of Maryland School of Medicine, 10 South Pine Street, MSTF 7th floor, Baltimore, MD 21201-1192, USA.  
 E-mail: fjiang@som.umaryland.edu

Received 27 July 2005; revised 19 October 2005; accepted 27 October 2005; published online 12 December 2005

In this study, we used a comprehensive approach that integrated simultaneous comparative genomic hybridization (CGH) and transcript microarray with proteomic analyses of six lung adenocarcinoma cell lines. We directly and specifically identified four putative oncogenes that could have been activated through amplification and consequent elevation of transcript expression. We used small interfering RNA (siRNA) to inhibit the expression of two of these four genes in the lung cancer cell lines, which further implicated them in oncogenesis. We then explored the clinical significance of these findings by assessing the expression of these two genes in tissue microarrays of human lung cancer specimens. Our findings underscore the power of integrated functional genomic analyses for identifying putative oncogenes in tumorigenesis; such activated genes could be useful as targets for diagnosis or therapy in lung cancer.

## Results

### *Simultaneous global genomic and transcript analyses identify 183 genes with increases in genomic copy numbers and transcript expression levels*

To identify genes in which increased DNA copy number might contribute to increased transcript in lung adenocarcinomas, first we used CGH with microarrays of six lung adenocarcinoma cell lines. We identified 587 genes showing increases in DNA copy number across all six cell lines (Supplementary Table 1S), which were distributed as 90 amplicons on all chromosomes except for chromosomes 13 and Y (Supplementary Table 2S). A subsequent transcript test with the identical arrays of the same cell lines revealed 275 genes that showed increased mRNA levels (Supplementary Table 3S). Using random permutation tests across all cancer cell lines, we identified 183 genes (31%) that showed elevated transcript levels from the 587 genes that were over-represented in the genome (Table 1), suggesting that elevated transcript levels of the 183 genes may reflect their genomic over-representation in the cancer cells. These findings are consistent with previous reports linking genomic changes with altered transcript patterns in breast cancer (Hyman *et al.*, 2002; Pollack *et al.*, 2002). However, our finding that only 31% of the genes showing increased DNA copy numbers had cognate increases in transcript expression in lung adenocarcinomas is different from the overall rates of 40–60% reported for breast cancer (Hyman *et al.*, 2002; Pollack *et al.*, 2002). This discordance may reflect methodologic differences between studies or biological differences between breast cancer and lung adenocarcinoma.

### *Proteomic analyses identify four genes for which protein abundance was associated with increases in the cognate gene and transcript levels*

Analysis of transcript patterns is insufficient for understanding the expression of protein products and the effect of genomic over-representation on the expression

of their cognate proteins. To extend these findings beyond genomic over-representation to expression of the protein products of those genes, we next assessed protein expression in the same cell lines by two-dimensional polyacrylamide gel electrophoresis (PAGE) and found that 42 different proteins, representing 42 individual genes, were significantly increased in the cancer cell lines (Table 2; Supplementary Figures 1S and 2S). Some of these proteins were identified as having multiple isoforms, and all individual isoforms exhibited increases in expression ranging from 4.6 to 12.8 times their expression in normal lung tissue cells. In comparing protein level of the 42 genes with changes in their cognate genomic and mRNA expression from the global microarray analyses, we found that four (9.5%) of those 42 genes – *PRDX1*, *EEF1A2*, *CALR*, and *KCIP-1* – showed statistically significant correlations between elevated protein expression and increases in both copy number and mRNA expression (all  $r > 0.84$ ;  $P < 0.05$ ) (Table 2) in the cancer cell lines. These findings imply that the abundance of these four proteins is attributable to the amplification and consequent elevated transcription of their cognate genes.

### *Validation of copy number, transcript, and protein expression of PRDX1, EEF1A2, CALR, and KCIP-1 in lung cancer cell lines*

To confirm our findings from the high-throughput analyses, we next used Southern, Northern, and Western blotting to assess DNA, RNA, and protein levels for the four genes identified in the six cell lines. For comparison, we arbitrarily chose one gene, *NFKB1*, in which an increase in protein level did not correlate with genetic changes. Overall, we found excellent concordance between the CGH microarray and Southern blotting analyses, transcript array and Northern blotting analyses, and proteomic and Western blotting analyses for all five genes (Figure 1). For example, *KCIP-1* showed fivefold amplification in five of the six cancer cell lines, whereas *NFKB1* showed no such increase in any of the cell lines. As for transcript expression, Northern blotting of *EEF1A2* showed high expression in five of the six cancer cell lines; again, levels of *NFKB1* transcript were not increased in any cancer cell line as compared with normal bronchial epithelial cells. The results of Western blotting were also consistent with the results of the proteomic experiments; for example, five of the cancer cell lines exhibited strong protein bands for *PRDX1* as compared with normal cells. These findings provide strong support for the validity of the results derived from the high-throughput techniques in this study.

These parallel analyses also revealed close correlations in the extent of changes in gene copies, transcript, and protein of each of the four genes in the cancer cell lines. For example, in the five cancer cell lines that showed at least fourfold increases in *EEF1A2* copy number, expression of transcript and protein was also increased by at least a factor of four as well (relative to their expression in normal cells) (Supplementary Figure 3S). The protein abundance of the four genes showing

Table 1 List of 183 genes with statistically significant correlation (0.05) between genomic copy number and transcript level

Gene symbol	Chro.	Distance from p arm of each chromosome (Mb)	$\alpha$
ENO1	1	8.5	0.0085
DDOST	1	20.1	0.0111
SFN	1	26.4	0.0113
MLP	1	32.2	0.0114
AKR1A1	1	45.4	0.0128
PRDX1	1	45.4	0.0122
UQCRH	1	46.2	0.0125
RPL7	1	96.4	0.0127
COL11A1	1	102.6	0.0129
MCL1	1	147.3	0.0222
PSMB4	1	148.1	0.0131
JTB	1	150.7	0.0134
RPS27	1	150.7	0.0135
HAX1	1	151	0.0266
MUC1	1	151.9	0.0143
CCT3	1	153.1	0.0167
CRABP2	1	153.4	0.0148
TKT	1	159.3	0.0152
ATP1B1	1	165.8	0.0234
CHIT1	1	199.7	0.0154
SNRPE	1	200.2	0.0165
YWHAQ	2	9.6	0.0159
ODC1	2	10.60	0.0119
RPL31	2	101.20	0.0161
BENE	2	110.40	0.0169
STAT1	2	191.80	0.0175
HSPD1	2	198.30	0.0277
HSPE1	2	198.30	0.0185
RPL37A	2	217.30	0.0388
IGFBP2	2	217.50	0.0189
RPS7	2	3.30	0.0193
RAB1A	2	65.30	0.0204
IGKC	2	89.00	0.0285
UTF	3	46.3	0.0455
PFN2	3	151	0.0207
KPNA4	3	161.5	0.0211
SI00P	4	6.7	0.1122
UGDH	4	39.3	0.0215
UCHL1	4	41.1	0.0222
SPPI	4	89.3	0.0227
TRIM2	4	154.7	0.0231
FGB	4	156	0.0235
FGG	4	156	0.0441
SDHA	5	0.251	0.0243
PDCD6	5	0.305	0.0245
CCT5	5	10.3	0.0446
PTPRF	5	14.2	0.0248
RPL37	5	40.8	0.0251
ENC1	5	74	0.0336
QP-C	5	132.2	0.0466
SPINK1	5	147.2	0.0256
CANX	5	179.2	0.0263
SOX4	6	21.7	0.0321
HDGF	6	22.6	0.0362
RPS10	6	34.6	0.0177
RPL10A	6	35.4	0.0369
VEGF	6	43.7	0.0372
OSF-2	6	45.4	0.0173
FSCN1	7	5.3	0.0378
CYCS	7	24.9	0.0381
CBX3	7	25.9	0.0289
IGFBP3	7	45.7	0.0389
CLDN4	7	72.7	0.0403
HSPB1	7	75.5	0.0433
CALR	7	92.7	0.0425
COL1A2	7	93.6	0.0457
ATPS12	7	98.7	0.0475
AKR1B10	7	133.6	0.0481

Table 1 (continued)

Gene symbol	Chro.	Distance from p arm of each chromosome (Mb)	$\alpha$
RPS20	8	56.7	0.0482
TCEB1	8	74.6	0.0486
LAPTM4B	8	98.5	0.0497
RPL30	8	98.7	0.0054
KCIP-1	8	101.6	0.0093
PABPC1	8	101.78	0.0119
EEF1D	8	144.4	0.0121
TSTA3	8	144.5	0.0122
RPL8	8	145.6	0.0128
TRA1	9	117.1	0.0136
RPL35	9	121.1	0.0133
HSPA5	9	121.5	0.0135
LCN2	9	124.4	0.0137
DPP7	9	133.4	0.0139
PFKP	10	3.2	0.0223
AKR1C1	10	5.1	0.0146
PLAU	10	75.6	0.0356
DSP	10	76.7	0.0289
TALDO1	11	0.434	0.0143
SLC22A1L	11	2.9	0.0151
TSSC3	11	2.9	0.0611
RPL27A	11	8.7	0.0156
ST5	11	8.8	0.0162
LDHA	11	18.5	0.0168
MDK	11	46.4	0.0162
DOC-1R	11	67.5	0.0167
MMP12	11	102.8	0.0177
HYOU1	11	118.9	0.0183
SCNN1A	12	6.3	0.0185
LDHB	12	21.7	0.0193
KRT7	12	52.3	0.0196
KRT5	12	52.6	0.0197
KRT6E	12	52.6	0.0201
ERBB3	12	56.2	0.0212
NACA	12	56.8	0.0218
TM4SF3	12	71.2	0.0401
NTS	12	86.2	0.0215
ASCL1	12	103.3	0.0219
TXNRD1	12	104.6	0.0223
CKAP4	12	106.6	0.0124
COX6A1	12	120.7	0.0435
BGN	12	122.5	0.0235
RAN	12	129.88	0.0238
RPL36A	14	48.1	0.0243
PGD	14	50.7	0.0248
THBS2	15	37.5	0.0251
TRAF4	15	38.3	0.0253
SPINT1	15	38.7	0.0254
RPL17	15	45.26	0.0411
PKM2	15	70.1	0.0258
IDH2	15	88.2	0.0211
RPL23A	16	0.377	0.0264
MSLN	16	0.753	0.0366
UBE2I	16	1.3	0.0271
RPS2	16	1.95	0.0281
CLDN9	16	3.1	0.0329
ARL6IP	16	18.7	0.0412
EIF3S8	16	28.3	0.0336
TUFM	16	28.9	0.0377
ALDOA	16	30.1	0.038
NME4	16	53.6	0.0381
GPR56	16	57.4	0.0386
CDH1	16	68.5	0.0289
NQO1	16	69.5	0.0396
SLC7A5	16	87.6	0.0397
APRT	16	88.6	0.0411
GALNS	16	88.6	0.0255
RPL13	16	89.3	0.0431
MCP	17	32.4	0.0465

Table 1 (continued)

Gene symbol	Chro.	Distance from p arm of each chromosome (Mb)	$\alpha$
ERBB2	17	35.11	0.0483
JUP	17	39.8	0.0495
CRF	17	40.39	0.0505
RPL27	17	41.1	0.0046
NME1	17	46.59	0.0082
COL1A1	17	48.6	0.0108
ABCC3	17	49.1	0.0326
NME2	17	49.6	0.0111
RPL38	17	72.7	0.0117
SMT3H2	17	73.6	0.0119
SYNGR2	17	76.6	0.0122
LGALS3BP	17	77.4	0.0127
P4HB	17	80.3	0.0126
PPAP2C	19	0.221	0.0228
GPI	19	39.55	0.0145
HPN	19	40.2	0.0129
ZNF146	19	41.4	0.0131
SPINT2	19	43.4	0.0238
PSMD8	19	43.5	0.0132
YIFIP	19	43.5	0.0135
RPS16	19	44.6	0.0144
CEACAM5	19	46.9	0.0145
CEACAM6	19	46.9	0.0143
GIPR	19	50.8	0.0259
SNRPD2	19	50.9	0.0413
KDELRL1	19	53.6	0.0152
RPL28	19	60.6	0.0156
RPS5	19	63.6	0.0267
TRIM28	19	63.7	0.0158
DAP	20	35.6	0.0166
TOP1	20	40.3	0.0172
UBE2C	20	45.1	0.0174
RPS21	20	61.6	0.0268
EEF1A2	20	62.8	0.0185
TFF3	21	42.6	0.0186
TFF1	21	42.7	0.0192
CSTB	21	44.1	0.0201
MIF	22	22.6	0.0202
XBPI	22	27.5	0.0204
PRDX4	X	22.9	0.0198
SYN1	X	46.3	0.0204
TIMP1	X	46.3	0.0209
PLP2	X	47.8	0.0212
MAGED1	X	50.3	0.0331
RPS4X	X	71	0.0124
SSR4	X	152.6	0.0232

corresponding increases in both DNA copy number and mRNA provides further evidence that these could be oncogenes, the activation of which is reflected by genomic amplification and consequent increases in transcript level in lung adenocarcinoma cell lines.

*Specific inhibition of EEF1A2 and KCIP-1 expression by siRNAs led to decreased cell proliferation and induction of apoptosis*

To further prove the oncogenic function of the identified genes in lung tumorigenesis, we used siRNAs to inhibit the endogenous expression of EEF1A2 and KCIP-1 protein in four lung cancer cell lines (H1563, H229, H522, and SK-LU). Transfection of the cancer cells with specific siRNAs reduced the level of EEF1A2 and KCIP-1 protein by 70–90% 48 h after transfection

(Supplementary Figure 4S). In contrast, EEF1A2 and KCIP-1 protein levels remained unchanged in mock-treated control cells and in cells transfected with a scrambled siRNA sequence. At 48 h after siRNA transfection, the percentage of proliferation of the transfected cancer cells was reduced to 15–30% as compared with 91–100% of cell proliferation of the same cell lines treated with PBS or scrambled siRNA (Supplementary Figure 5S). Apoptosis of siRNA-transfected cells was 27–34%, whereas only 4% of the same cell lines treated with PBS or scrambled siRNA showed apoptosis. These results strongly support an oncogenic role for the identified genes in lung cancer and confirm their potential usefulness as therapeutic targets for the disease.

*Amplification and protein expression of KCIP-1 and EEF1A2 in lung tissue*

To further validate these findings and to assess the possible clinical significance of the four potential putative oncogenes identified from the cell lines, we first applied fluorescence *in situ* hybridization and immunohistochemical analysis, in parallel, to commercially available human lung tissue microarrays (Ambion, Austin, TX, USA) to evaluate the status of two of these four genes in lung cancer tissue specimens. (Commercially available antibodies to PRDX1 or CALR were not suitable for use in immunohistochemical analysis when this report was written.) Overexpression of KCIP-1 and EEF1A2 protein in the tumors was concordant with amplification of the corresponding genes ( $P=0.0003$  for KCIP-1 and  $P=0.0011$  for EEF1A2). For example, 16 (35%) of the 46 lung adenocarcinomas in the microarray showed amplification of *KCIP-1*, and strong cytoplasmic staining for KCIP-1 protein was seen in 18 tumors (39%) (Figure 2). We next examined whether overexpression of these genes was associated with increased cell proliferation by analysing Ki-67 expression in contiguous sections of the tissue microarrays. Positive Ki-67 expression was found to correlate with positive expression of both KCIP-1 ( $P=0.02$ ) and EEF1A2 ( $P=0.01$ ). To extend these findings, we then studied 11 tissue microarray blocks comprising normal and tumor tissue specimens from 113 patients with pathologic stage I non-small-cell lung cancer who had undergone curative surgery (Wang *et al.*, 2005). Immunohistochemical analysis showed that EEF1A2 was expressed in 32 cases (28%) and KCIP-1 in 29 cases (26%). Univariate and multivariate Cox proportional hazards models were used to detect possible associations between EEF1A2 and KCIP-1 expression and clinicopathologic variables. Expression of EEF1A2 or KCIP-1 was associated with short overall survival time ( $P=0.0012$  for EEF1A2 and  $P=0.0026$  for KCIP-1) (Supplementary Figure 6S). Age at diagnosis, histologic type of cancer, degree of tumor differentiation, and smoking history were not associated with survival time.

Although only two genes were validated in the lung tissue microarrays (because available antibodies to the other two genes were not suitable for use in

**Table 2** Proteins showing significant overexpression in cancer cell lines relative to those in normal bronchial epithelial cell lines and their correlation coefficients with increased DNA copy number or mRNA values\*

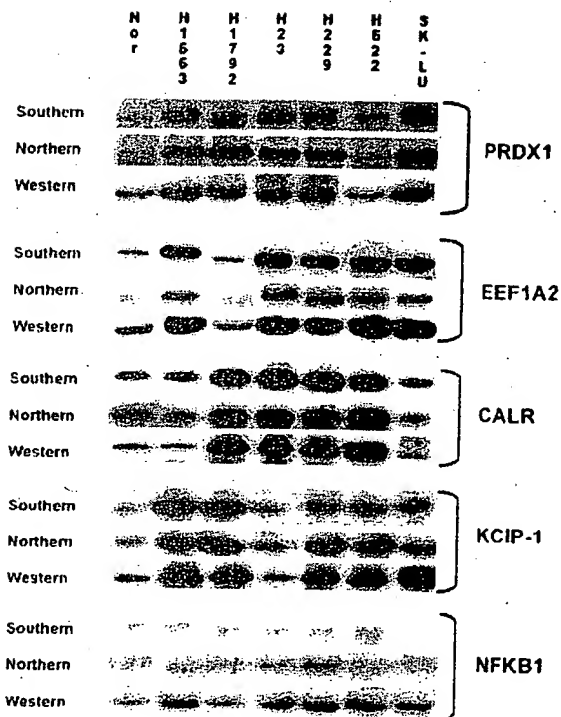
Acc. no.	Gene ID	Gene	Mw/pI	Description	r with genomic copy changes <sup>a</sup>	r with mRNA changes <sup>a</sup>
Q06830	5052	PRDX1	48.4/5.4	Peroxisiredoxin 1	<b>0.92364</b>	<b>0.91892</b>
Q05639	1917	EEF1A2	50.5/5.7	Eukaryotic translation elongation factor 1 alpha 2	<b>0.90218</b>	<b>0.89456</b>
P27797	811	CALR	61/5.5	Calreticulin	<b>0.84128</b>	<b>0.86434</b>
P63104	7534	KCIP-1	27/6.5	Tyrosine 3-monooxygenase activation protein, zeta	<b>0.84467</b>	<b>0.85499</b>
P07237	5034	P4HB	54/6.2	Procollagen-proline, 2-oxoglutarate 4-dioxygenase	<b>0.91884</b>	<b>0.76786</b>
Q04695	3872	KRT17	48.0/4.9	Keratin 17	<b>0.00236</b>	<b>0.86892</b>
P09211	2950	GSTP1	23.2/4.7	Glutathione S-transferase pi	<b>0.84218</b>	<b>0.69456</b>
P17936	3486	IGFBP-3	31.6/5.8	Insulin-like growth-factor binding protein 3	<b>0.06412</b>	<b>0.16434</b>
P26641	1937	EEF1G	50/6.4	Eukaryotic translation elongation factor 1 gamma	<b>0.00446</b>	<b>0.85549</b>
P08727	3880	KRT19	44.1/5.2	Keratin 19	-0.04884	<b>0.86786</b>
P04792	3315	HSPB1	22/6.5	Heat shock 27 kDa protein 1	<b>0.00364</b>	<b>0.31892</b>
P00558	5230	PGK1	44.5/4.2	Phosphoglycerate kinase 1	<b>0.50402</b>	<b>0.79456</b>
Q01995	5876	TAGLN	22.5/4.3	Transgelin	-0.34128	-0.26434
P08631	3055	JTK9	59.5/6.8	Hemopoietic cell kinase	-0.01446	<b>0.02549</b>
P09382	3956	LGALS1	16/5.5	Galectin-1, galactoside-binding, soluble, 1	<b>0.026623</b>	<b>0.01123</b>
Q92784	8110	DPF3	25.8/4.8	D4, zinc and double PHD fingers, family 3	<b>0.094884</b>	-0.03214
P54257	9001	HAP1	75.5/6.5	Huntington-associated protein 1	<b>0.12364</b>	-0.08108
P05783	3875	KRT18	48/5.3	Keratin 18	<b>0.010218</b>	<b>0.60544</b>
P05787	3856	KRT8	9.2/4.4	Keratin 8	<b>0.041280</b>	<b>0.84566</b>
P00738	3240	HP	55.2/6.2	Haptoglobin	<b>0.044679</b>	-0.14501
P09769	2268	FGR	59.5/5.2	Gardner-Rasheed feline sarcoma viral oncogene homolog	<b>0.031264</b>	-0.13789
P19838	4790	NFKB1	50.4/6.3	Nuclear factor of kappa light gene enhancer in B-cells 1	<b>0.04467</b>	-0.14501
P29034	6273	S100A2	10.9/4.6	S100 calcium-binding protein A2	<b>0.87964</b>	<b>0.243214</b>
Q13105	7709	ZBTB17	87.9/5.3	Zinc-finger and BTB domain containing 17	-0.17636	<b>0.048108</b>
Q00987	4193	MDM2	75.2/4.8	Transformed 3T3 cell double minute 2	-0.19782	-0.50544
P27816	4134	MAP4	111/5.4	Microtubule-associated protein 4	<b>0.25872</b>	-0.05356
P52732	3832	KIF11	119.2/6.2	Kinesin family member 11	-0.25778	-0.53444
P25205	4172	MCM3	90.9/5.5	Minichromosome maintenance deficient 3	<b>0.25644</b>	<b>0.053666</b>
P08631	3055	HCK	59.5/5.7	Hemopoietic cell kinase	<b>0.65533</b>	<b>0.054501</b>
P09237	4316	MMP7	22.6/5.8	Matrix metalloproteinase 7	<b>0.234987</b>	<b>0.876820</b>
P30305	994	CDC25B	64.9/4.5	Cell division cycle 25B	<b>0.045116</b>	<b>0.283214</b>
P50290	998	CDC42	21.3/6.1	Cell division cycle 42 (GTP-binding protein, 25 kDa)	-0.47636	<b>0.088108</b>
P61586	387	RHOA	19.8/6.9	Ras homolog gene family, member A	-0.49782	-0.00544
P63000	5879	RAC1	21.5/6.8	Ras-related C3 botulinum toxin substrate 1	-0.05583	-0.03566
P07437	203068	TUBB	49.6/6.5	Tubulin, beta polypeptide	<b>0.255533</b>	<b>0.145010</b>
P24864	898	CCNE1	47.1/4.3	Cyclin E1	-0.65116	<b>0.232149</b>
P04141	1437	CSF2	16.9/6.3	Colony stimulating factor 2 (granulocyte-macrophage)	-0.64636	-1.28108
P28072	5694	PSMB6	25.3/5.2	Proteasome (prosome, macropain) subunit, beta type, 6	-0.69782	-1.30544
P00352	216	ALDH1A1	54.7/4.3	Aldehyde dehydrogenase 1 family, member A1	-0.75872	<b>0.03356</b>
Q03013	2948	GTM4	25.3/5.0	Glutathione S-transferase M4	-0.78533	<b>0.134501</b>
P63241	1984	EIF5A	10/4.4	Eukaryotic translation initiation factor 5A	-0.97893	-1.44321
Q01469	2171	EFABP	18.0/4.2	Fatty acid-binding protein 5	<b>0.25684</b>	-0.36432

\*Only the gene showing statistically significant increased protein expression with increases in both genomic copy number and transcript simultaneously will be considered as potential putative oncogene in lung adenocarcinoma cells. <sup>a</sup>r, Spearman correlation coefficients between proteins and genomic or mRNA values are based on all six cancer cell lines; bold indicates  $P < 0.05$ , if  $r > 0.84000$ . Mw, molecular weight; pI, isoelectric point.

immunohistochemical analysis), these findings are consistent with those from our cell lines, demonstrating again that genomic amplification and consequent increases in amounts of transcript may be, at least in part, driving the abundance of proteins in these lung tumors. The association between expression of these genes and that of Ki-67, a known indicator of poor prognosis in lung cancer (Martin *et al.*, 2004), suggests that activation of these genes may be an indicator of tumor aggressiveness. These results also suggest that expression of EEF1A2 and KCIP-1 proteins in stage I non-small-cell lung cancer may be useful as a marker for distinguishing patients with relatively poor prognosis from those who might benefit from adjuvant treatment.

## Discussion

Our current study illustrates the power of integrated functional genomic analyses for identifying putative oncogenes and for evaluating their potential clinical significance. Among the four identified oncogenes, three genes (*PRDX1*, *CALR*, and *KCIP-1*) have been implicated in lung tumorigenesis. *PRDX1* is an antioxidant protein involved in regulating cell proliferation, differentiation, and apoptosis. Kim *et al.* (2003) found *PRDX1* expression to be elevated in both lung cancer and adjacent normal lung tissue, suggesting that activation of *PRDX1* may enhance proliferation in lung cancer. *CALR* has a major role in  $\text{Ca}^{2+}$  binding and the



**Figure 1** Confirmation by Southern, Northern, and Western blot analyses of increased DNA copies, transcript levels, and protein levels in the four genes identified in high-throughput analyses. For comparison, we arbitrarily chose one gene, *NFKB1*, in which an increased protein level did not correlate with genetic changes. The blotting results are consistent with the results from the CGH array, transcript array, and proteomic analyses. Nor, indicates normal bronchial epithelial cell line. All the experiments were repeated at least three times with each cell line. Means of normalized to  $\beta$ -actin signal intensities on Southern, Northern, and Western blots, along with 95% confidence intervals, were calculated ( $\beta$ -actin signals are not shown in the figure; two different normal bronchial epithelial cell lines were used in the confirmation and only one normal cell line is shown in the figure).

transcriptional regulation of other genes and was recently found to be overexpressed in 73% of 40 lung adenocarcinomas (Oates and Edwards, 2000). *KCIP-1* belongs to the 14-3-3 family, which participates via the MAPK and Wnt signaling pathways in the regulation of many cellular processes including cell proliferation and differentiation as well as tumorigenesis (Thomas *et al.*, 2005). *KCIP-1* was recently found to be expressed in all 12 lung tumors tested in a single-institution study (Qi *et al.*, 2005). Interestingly, *EEF1A2* was originally considered a putative oncogene in ovarian cancer on the basis of its being amplified in 25% and overexpressed in 30% of the same set of ovarian tumors (Anand *et al.*, 2002); functional analyses have established its oncogenic role in cellular transformation (Lee, 2003). Our discovery that *EEF1A2* may be a putative oncogene in lung adenocarcinoma demonstrates the power of our functional genomic strategy for rapidly identifying potential oncogenes.

Although the main focus of this study was to specifically identify putative oncogenes, it should be

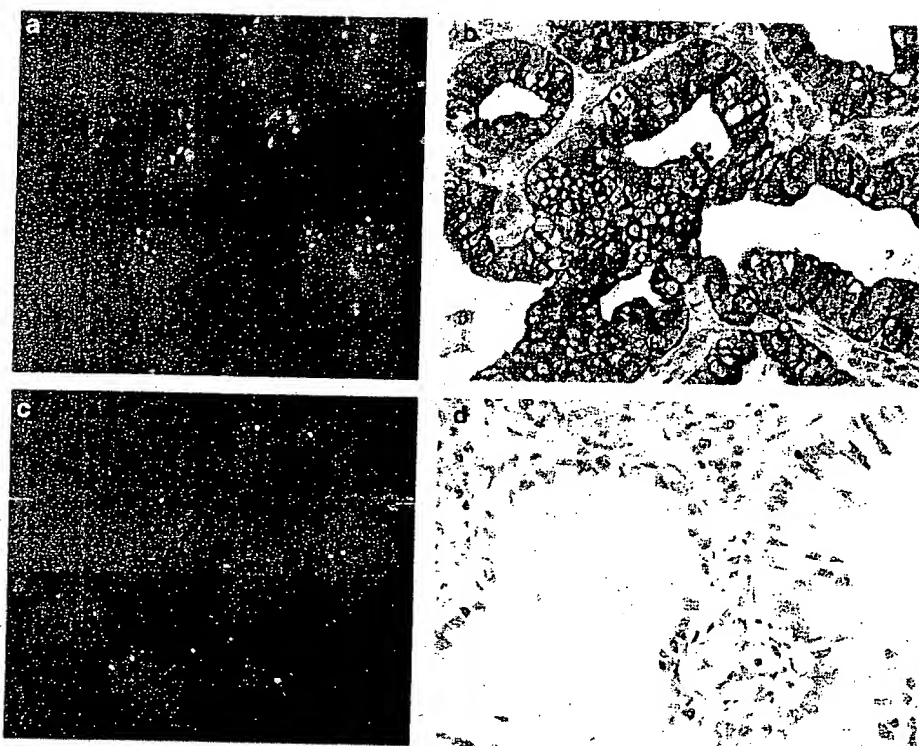
noted that 90.7% of the genes showing high protein expression did not show corresponding increases in both DNA copy number and transcript, a finding consistent with that of others that transcriptional, translational, and post-translational regulatory mechanisms can greatly influence the abundance of protein in lung tumorigenesis (Chen *et al.*, 2002). For example, *NFKB1* is a critical arbiter of immune responses, cell survival, and transformation and is often activated in several types of tumors (Chen *et al.*, 2002). Deregulation of *NFKB1* is thought to be modulated through phosphorylation of Ser337 by protein kinase A (Chen *et al.*, 2002). In our study, 68.8% of the genes showing over-representation in the genome did not show elevated transcript levels, implying that at least some of these genes are 'passenger' genes that are concurrently amplified because of their location with respect to amplicons but lack biological relevance in terms of the development of lung adenocarcinoma.

Although the potential oncogenes we identified here are likely to be important, certainly other oncogenes could be involved in the development of lung adenocarcinoma. The oligo microarray we used consists of 22 000 probes, which represent only about 60% of the human genome. Moreover, each probe was designed for the 3' region of expressed sequence tags of the selected genes. Also, our results were initially derived from cancer cell lines, although the findings were later confirmed in human tissue samples. Our ongoing study using microarrays with information on more genes and the development of high-resolution proteomic analyses for use with larger numbers of specimens will allow more comprehensive analyses of the molecular consequences of gene amplifications. Such expanded analyses will very likely lead to the identification of additional oncogenes.

Some of the results of our current study were comparable to those of other studies of lung cancer. For example, genomic copy number and protein levels of *KCIP-1* were previously found to be amplified and overexpressed in primary lung cancers by cDNA clone-based CGH array analysis (Jiang *et al.*, 2004) and proteomic analysis (Chen *et al.*, 2002), respectively. Our functional genomic approach, which integrates simultaneous CGH, transcript microarrays, proteomic analyses, and siRNA, allows us not only to quickly identify potential oncogenes but also to explore their significance as diagnostic and therapeutic targets in tumor progression – more than could be achieved by any technique alone.

Genes identified in this way may serve as promising targets for diagnosis and therapy in lung adenocarcinoma. Further research on the clinical implications of such genes is needed; experiments now underway in our laboratory include overexpression of the genes in normal cells, disruption of the function of these genes in cancer cells, and investigation of how interactions among these genes (or interactions with other known oncogenes) may mediate the expression of the transformed phenotype.





**Figure 2** *EEF1A2* amplification is associated with high *EEF1A2* protein expression in lung adenocarcinomas. (a) Cells from a lung adenocarcinoma sample in which *EEF1A2* is amplified show more green signals (*EEF1A2*) than red signals (chromosome 20 centromeric probe) (original magnification,  $\times 400$ ). (b) Immunohistochemical staining of cells from the same tissue sample as in panel a shows strong *EEF1A2* staining in the cytoplasm. (c) A lung adenocarcinoma sample with two copies of *EEF1A2* and chromosome 20 centromeric probe, indicating no *EEF1A2* amplification (original magnification,  $\times 400$ ). (d) Immunohistochemical staining of cells from the same tissue sample as in panel c shows negative staining for *EEF1A2*.

## Materials and methods

### Cell lines

Six human lung adenocarcinoma cell lines (H23, H229, H1792, SK-LU-1, H522, and H1563) were obtained from the American Type Culture Collection (Manassas, VA, USA). Two normal bronchial-epithelial cell lines were obtained from Clontech (Palo Alto, CA, USA). Genomic DNA, mRNA, and protein were derived from a single harvest of these cells.

### DNA and RNA profiles by microarray analysis

Genomic DNA labeling and hybridization were performed as described previously (Barrett *et al.*, 2004) with Agilent's Human 1A Oligo Microarray (V2) (Agilent Technologies, Palo Alto, CA, USA), which contains 22,000 unique 60-mer oligos. Details of the protocol for analysing transcripts are available at <http://www.chem.agilent.com>. Map positions for arrayed genes were assigned by identifying the DNA sequence represented in the UniGene cluster and matching it with the Golden Path genome assembly (<http://genome.ucsf.edu/>; Mat 7, 2004 Freeze). Microarray images of DNA copy number and expression were analysed by using AgilentCGH Analytics and Feature Extraction software. DNA copy number profiles that deviated significantly from background signal ratios (measured from normal control cell hybridization, as described elsewhere; Barrett *et al.*, 2004) were interpreted as evidence of true differences in DNA copy number. The criteria for defining genomic over-representation and amplicons are described elsewhere (Hyman *et al.*, 2002); details are given in the

Supplementary Information. An increase in mRNA level was defined as a twofold increase in signal ratio relative to that of the control ( $\log_2 > 1$ ).

### Quantitative two-dimensional PAGE and mass spectrometry

Analysis of proteins by two-dimensional PAGE and their identification by mass spectrometry were performed as previously described (Shen *et al.*, 2004). Briefly, protein pellets were solubilized in rehydration buffer, after which the first-dimension isoelectric focusing was carried out with a Protean IEF Cell (Bio-Rad Laboratories) and the second-dimension separation was carried out with Bio-Rad's Ready Gel Precast Gels and the Bio-Rad Criterion Cell apparatus. Protein spots were visualized by silver-based staining, and all gels were assessed with Bio-Rad's PDQuest 2D gel image analysis software. Selected spots were subjected to in-gel tryptic digestion and analysed on a Voyager-DE PRO matrix-assisted laser desorption/ionization/time-of-flight mass spectrometer (Applied Biosystems, Foster City, CA, USA). The mass list of the 20 most intense monoisotopic peaks for each sample was entered in the MS-Fit search program (v3.2.1) (<http://prospector.ucsf.edu/ucsfhtml4.0/msfit.htm>) and searched in the National Center for Biotechnology Information protein database.

### Southern, Northern, and Western blot analyses

Southern, Northern, and Western blot hybridizations were performed according to standard protocols. cDNA clones for the tested genes were purchased from Invitrogen (Carlsbad,



CA, USA) and prepared as probes for the blot hybridizations. Antibodies used were obtained as follows: PRDX1, CALR, NFKB1, KCIP-1, and  $\beta$ -actin from Santa Cruz Biotechnology (Santa Cruz, CA, USA); and EEFA2 from Upstate Biotechnology (Waltham, MA, USA).

*Fluorescence in situ hybridization and immunohistochemical analyses of lung tissue microarrays*

Fluorescence *in situ* hybridizations and immunohistochemical analyses of KCIP-1 and EEFA2 were carried out as described elsewhere (Jiang et al., 2002; Wang et al., 2005) with Lung Tissue Microarrays (Ambion, Austin, TX, USA) and 11 homemade microarray blocks containing tissue samples from 113 patients with pathologic stage I non-small-cell lung cancer (Wang et al., 2005). DNA probes specific for KCIP-1 and EEFA2 were obtained by screening a Human BAC Clone library (Invitrogen) by polymerase chain reaction as described previously (Jiang et al., 2002). The antibodies used for the immunohistochemical analyses were the same as those used for the Western blotting. Cell proliferation of the lung tissues was assessed with a Ki-67 monoclonal antibody from Santa Cruz Biotechnology. Definitions of the cutoff value for a positive result of each antibody are shown in Supplementary Information.

*siRNA transfection, cellular proliferation assay, and apoptosis analysis*

Transfections were carried out by using siPORT Lipid Transfection Agent (Ambion) with siRNAs targeting KCIP-1 or EEFA2 or with a scrambled siRNA duplex (siControl) (Dharmacon Inc., Lafayette, CO, USA), with PBS used as a negative control (Jiang et al., 2002). Cells were fixed 24, 48, or 96 h later and subjected to further tests. All siRNAs were prepared by using a transcription-based method with Silencer siRNA according to the manufacturer's instructions (Ambion). Sequences of the individual siRNAs are listed in Supplementary Table 4S. Inhibition of cell growth by the

siRNAs was determined by MTT staining, and cell growth rate was plotted against the percentage of viable cells in the saline-treated controls (a value arbitrarily set at 100%) (Jiang et al., 2002). Apoptosis was analysed by fluorescence cell cycle analysis of terminal deoxynucleotidyl transferase-mediated dUTP nick-end labeling with FITC-labeled dUTP (Boehringer Mannheim Biochemicals, Mannheim, Germany) (Jiang et al., 2005).

*Statistical analyses*

Relationships between gene copy number and mRNA level were examined as described elsewhere (Hyman et al., 2002, Supplementary Information). Correlations between protein abundance and DNA copy number and mRNA expression of the corresponding genes were evaluated with the Spearman correlation coefficient. Fisher's exact test and  $\chi^2$ -tests were used to analyse associations between amplification and expression of the candidate genes with various histopathologic variables of the samples in the tissue microarrays. Univariate and multivariate analyses were carried out with Cox's proportional hazards model to determine which independent factors might have a joint significant influence on survival. A  $P$ -value  $\leq 0.05$  was considered statistically significant; all statistical tests were based on a two-sided significance level.

**Acknowledgements**

This work was supported by National Institutes of Health Grant CA113707-01, an institutional research grant from The University of Texas MD Anderson Cancer Center, a Developmental Project/Career Development Award from The University of Texas Specialized Programs of Research Excellence in Lung Cancer P50 CA70907, and an M Keck Center for Cancer Gene Therapy Award (all to FJ). We thank Christine F Wogan of the Department of Scientific Publications for editorial review of this manuscript.

**References**

- Anand N, Murthy S, Amann G, Wernick M, Porter LA, Cukier IH et al. (2002). *Nat Genet* 31: 301–305.
- Barrett MT, Scheffer A, Ben-Dor A, Sampas N, Lipson D, Kincard R et al. (2004). *Proc Natl Acad Sci USA* 101: 17765–17770.
- Chen G, Gharib TG, Huang CC, Taylor JM, Misek DE, Kardis SL et al. (2002). *Mol Cell Proteomics* 1: 304–313.
- Hyman E, Kauraniemi P, Hautaniemi S, Wolf M, Mousset S, Rozenblum E et al. (2002). *Cancer Res* 62: 6240–6245.
- Jiang F, Caraway NP, Li RY, Katz RL. (2005). *Oncogene* 24: 3409–3418.
- Jiang F, Lin F, Price R, Gu J, Medeiros LJ, Zhang HZ et al. (2002). *J Mol Diagn* 4: 144–149.
- Jiang F, Yin Z, Caraway NP, Li R, Katz RL. (2004). *Neoplasia* (New York) 6: 623–635.
- Kim HJ, Chae HZ, Kim YJ, Kim YH, Hwang TS, Park EM et al. (2003). *Cell Biol Toxicol* 19: 285–298.
- Lee JM. (2003). *Reprod Biol Endocrinol* 1: 69–73.
- Luk C, Tsao MS, Bayani J, Shepherd F, Squire JA. (2001). *Cancer Genet Cytogenet* 125: 87–99.
- Martin B, Paesmans M, Mascaux C, Berghmans T, Lohaire P, Meert AP et al. (2004). *Br J Cancer* 91: 2018–2225.
- Oates J, Edwards C. (2000). *Histopathology* 36: 341–347.
- Pollack JR, Sorlie T, Perou CM, Rees CA, Jeffrey SS, Lonning PE et al. (2002). *Proc Natl Acad Sci USA* 99: 12963–12968.
- Qi W, Liu X, Qiao D, Martinez JD. (2005). *Int J Cancer* 113: 359–363.
- Ross JS, Fletcher JA. (1999). *Semin Cancer Biol* 9: 125–138.
- Shen J, Person MD, Zhu J, Abbuzzese JL, Li D. (2004). *Cancer Res* 64: 9018–9026.
- Thomas D, Guthridge M, Woodcock J, Lopez A. (2005). *Curr Top Dev Biol* 67: 285–303.
- Tonon G, Wong KK, Maulik G, Brennan C, Feng B, Zhang Y et al. (2005). *Proc Natl Acad Sci USA* 102: 9625–9630.
- Wang H, Zhang Z, Li R, Ang KK, Zhang H, Caraway NP et al. (2005). *Int J Cancer* 116: 285–290.

Supplementary Information accompanies the paper on Oncogene website (<http://www.nature.com/onc>)